DOCUMENT RESUME

ED 381 360 SE 056 089

TITLE	Improving Science and Mathematics Education: A Toolkit for Professional Developers: Alternative Assessment.
INSTITUTION	Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC. Office of Reform Assistance and Dissemination.
PUB DATE	Nov 94
NOTE	574p.; Prepared by the Regional Educational Laboratory Network Program on Science and Mathematics Alternative Assessment.
AVAILABLE FROM	Northwest Regional Educational Laboratory, 101 S.W. Main Street, Suite 500, Portland, OR 97204.
PUB TYPE	Guides - Classroom Use - Teaching Guides (For Teacher) (052)
EDRS PRICE	MF02/PC23 Plus Postage.
DESCRIPTORS	*Constructivism (Learning); Educational Change; Elementary Secondary Education; *Learning Activities; *Mathematics Instruction; Science Activities; *Science Instruction; *Student Evaluation
IDENTIFIERS	*Alternative Assessment; Mathematics Activities; *Reform Efforts

ABSTRACT

Since assessment is intimately linked with instruction, it too is changing under the wave of numerous efforts at reforming mathematics and science education. From this new perspective, the learner actively constructs personal meaning from information and experiences by linking new information with his or her pre-existing knowledge and understanding. This document was designed as a professional development resource to assist teachers in deepening their vision of what students need to know and be able to do, and to change the interactions between teachers and students in the classroom. It is a compilation of activities and supportive materials developed not only as an alternative assessment resource, but as a means for engaging teachers in dialogue about changing the ways mathematics and science have traditionally been taught and assessed. The Toolkit contains the following chapters: (1) "Introducing the Toolkit," (2) "Supporting Educational Improvement with Alternative Assessment," (3) "Integrating Assessment with Curriculum and Instruction," (4) "Exploring Design Options for Alternative Assessment," (5) "Evaluating the Quality and Equity of Alternative Assessments," (6) "Using Alternative Assessment in Grading and Reporting," and (7) "Designing Effective Professional Development." Appendices include alternative assessment samples and a resource listing of science and mathematics reform initiatives, agencies, state departments of education and school districts, publishers, print materials, and sources of information on tests. A 53-item annotated list of articles and books concludes the toolkit. (MKR)

ED 381 360

Improving Science and Mathematics Education

A Toolkit for Professional Developers: Alternative Assessment

November 1994

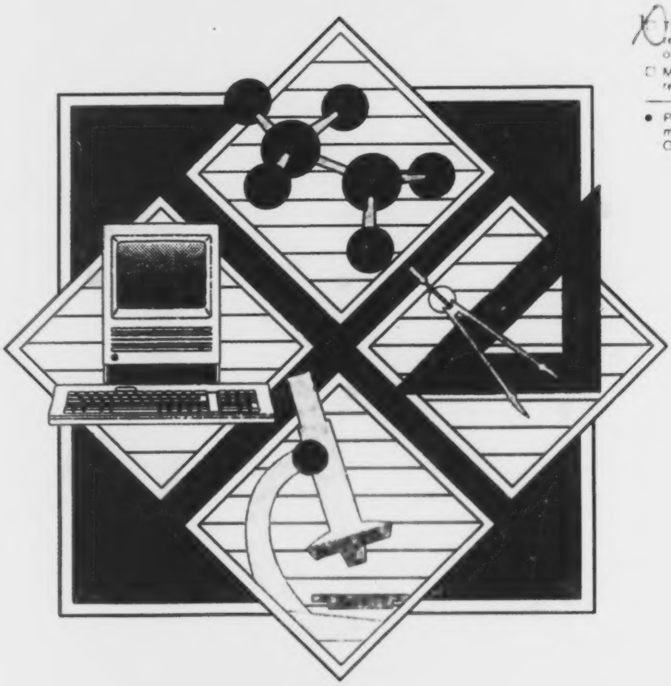
U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

---

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy



REGIONAL EDUCATIONAL
LABORATORY NETWORK

Sponsored by the U.S. Department of Education
Office of Educational Research and Improvement (OERI)

SE 056 039

BEST COPY AVAILABLE

REGIONAL EDUCATIONAL LABORATORIES

Appalachia Educational Laboratory
Charleston, West Virginia

Far West Laboratory for Educational Research and Development
San Francisco, California

Mid-continent Regional Educational Laboratory
Aurora, Colorado

North Central Regional Educational Laboratory
Oak Brook, Illinois

Northwest Regional Educational Laboratory
Portland, Oregon

Pacific Region Educational Laboratory
Honolulu, Hawaii

The Regional Laboratory for Educational Improvement
of the Northeast and Islands
Andover, Massachusetts

Research for Better Schools
Philadelphia, Pennsylvania

SouthEastern Regional Vision for Education
Greensboro, North Carolina

Southwest Educational Development Laboratory
Austin, Texas

This publication is sponsored wholly or in part by the U.S. Department of Education, Office of Educational Research and Improvement, under contract numbers: RP91002001 through RP91002010. The content of this publication does not necessarily reflect the views of any individual laboratory, the Department of Education, or any other agency of the U.S. Government. This publication is available from Northwest Regional Educational Laboratory.

© Copyright, Northwest Regional Educational Laboratory, 1994, all rights reserved

Improving Science and Mathematics Education

A Toolkit for Professional Developers: Alternative Assessment

November 1994

By:

Regional Educational Laboratory Network Program
on Science and Mathematics Alternative Assessment

Available from:

Northwest Regional Educational Laboratory
101 SW Main Street, Suite 500
Portland, Oregon 97204
(503) 275-9500; Fax: (503) 275-9489
Document Reproduction Service: (503) 275-9519

REGIONAL EDUCATIONAL
LABORATORY NETWORK

Sponsored by the U.S. Department of Education
Office of Educational Research and Improvement (OERI)

Acknowledgements

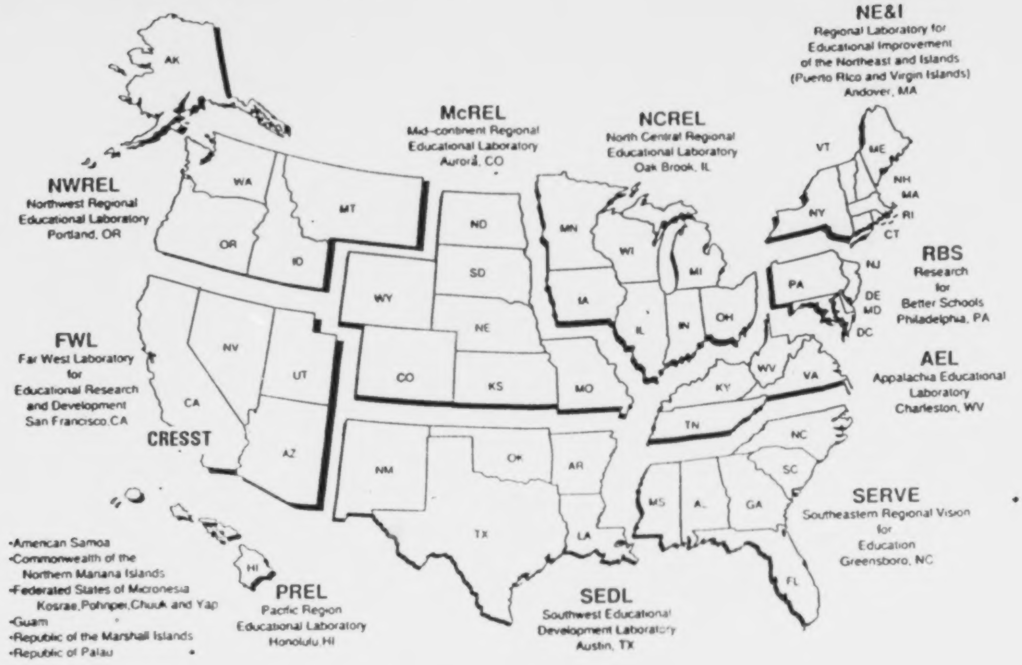
Many people contributed time and ideas during production of the *Toolkit*. We would especially like to acknowledge the following individuals.

Thanks to Jay Millman (Cornell University), Jay McTighe (Maryland Assessment Consortium), and Elliott Asp (Littleton School District, Colorado) for reviewing a 1993 draft of the *Toolkit* and providing ideas for revision.

Thanks to all those authors and organizations willing to share their efforts by allowing us to reprint portions of their alternative assessments: Alberta Education, Canada; California Department of Education; Center for Technology in Education, Bank St. College of Education; National Center for Research on Evaulation, Standards, and Student Testing (CRESST), University of California Los Angeles; Far West Eisenhower Regional Consortium; Kentucky Department of Education; Lake County Educational Service Center, Illinois; Leon Paulson, Multnomah Educational Service District, Oregon; Maryland Department of Education; Mathematical Sciences Education Board; Michael Lehman, Holt Public Schools, Michigan; New Standards Project; Oregon Department of Education; Richard Shavelson, University of California at Santa Barbara; Rodney Doran, State University of New York at Buffalo; Hawaii Algebra Learning Project, University of Hawaii; Utah Department of Education.

Thanks to those who helped with the pilot test and provided additional insights: Michael Power, Everett, Washington; Chestine Edgar, Tacoma, Washington; Art Schaefer, Ketchikan, Alaska; Tom Peters and the PASS grant teachers, Wichita Falls, Texas.

Without the leadership and support of the laboratory directors, this work would not have been possible.


NE&I
Regional Laboratory for Educational Improvement of the Northeast and Islands (Puerto Rico and Virgin Islands)
Andover, MA
McREL
Mid-continent Regional Educational Laboratory
Aurora, CO
NCREL
North Central Regional Educational Laboratory
Oak Brook, IL
NWREL
Northwest Regional Educational Laboratory
Portland, OR
RBS
Research for Better Schools
Philadelphia, PA
FWL
Far West Laboratory for Educational Research and Development
San Francisco, CA
AEL
Appalachia Educational Laboratory
Charleston, WV
CRESST
SERVE
Southeastern Regional Vision for Education
Greensboro, NC
•American Samoa
•Commonwealth of the Northern Mariana Islands
•Federated States of Micronesia
Kosrae, Pohnpei, Chuuk and Yap
•Guam
•Republic of the Marshall Islands
•Republic of Palau
PREL
Pacific Region Educational Laboratory
Honolulu, HI
SEDL
Southwest Educational Development Laboratory
Austin, TX
AK
WA
OR
ID
MT
ND
SD
WY
NE
CO
KS
NV
UT
CA
AZ
NM
OK
TX
MN
IA
MO
AR
LA
WI
IL
MI
IN
OH
KY
TN
MS
AL
GA
FL
SC
NC
VA
WV
PA
NY
VT
ME
NH
MA
RI
CT
NJ
DE
MD
DC
HI

Network Participants

Ullik Rouk
CEDaR
2000 L St., NW, Suite 601
Washington, D.C. 20036
(202) 223-1593

Sue Gruskin
OERI
555 New Jersey Ave., NW
Washington, D.C. 20208
(202) 219-2132

Carolyn Carter
Patricia Kusimo
AEL
PO Box 1348
Charleston, WV 25325
(304) 347-0470

Joan Herman
CRESST
405 Hilgard Ave.
145 Moore Hall
Los Angeles, CA 90024
(310) 206-1532

Ann Muench
Jerome Shaw
Gary Estes
FWL
730 Harrison St.
San Francisco, CA 94107
(415) 565-3000

Alice Krueger
McREL
2550 S. Parker Rd.
Suite 500
Aurora, CO 80014
(303) 337-0990

Linda Bond
Cathy Cook
Gil Valdez
NCREL
1900 Spring Rd
Suite 300
Oak Brook, IL 60521
(708) 571-4700 or
(317) 328-9704

Eileen Ferrance
The Regional Laboratory - Northeast and Islands
300 Brickstone Square
Suite 900
Andover, MA 01810
(508) 470-0098
(800) 347-4200

Judy Arter
Dean Arrasmith
NWREL
101 SW Main St.
Portland, OR 97204
(503) 275-9500

Kathy Busick
PREL
828 Fort Street Mall
Suite 500
Honolulu, HI 96813
(808) 533-6000

Janice Dilliplane-Kruse
Keith Kershner
RBS
444 N. Third St.
Philadelphia, PA 19123
(215) 574-9300, ext 279

Jackie Palmer
Glenda Clark
Wes Hoover
SEDL
211 E. Seventh St.
Austin, TX 78701
(512) 476-6861

Wendy McColskey
SERVE
P. O. Box 5367
Greensboro, NC 27435
(910) 334-3211
(800) 755-3277

Mahna Schwager
SWRL
4665 Lampson Ave.
Los Alamitos, CA 90720
(310) 598-7661

BEST COPY AVAILABLE

Authors

The following people contributed to sections of the *Toolkit*:

Judy Arter, NWREL	Lead lab; Chapters 1, 2, 3, 4, 5, 7; Activities 2.1, 2.5, 2.6, 2.7, 2.9, 2.10, 3.1, 4.1, 4.3, 5.1; Samples 4.1 through 4.10, 5.1 through 5.9
Linda Bond, NCREL	Chapter 2; Activity 2.8
Kathy Busick, PREL	Chapters 6, 7; Activities 2.3, 2.4, 2.5, 5.3, 5.5, 6.2
Carolyn Carter, AEL	Chapter 5; Activities 3.3, 5.2
Glenda Clark, SEDL	Activity 2.9
Melissa Chatpa, NCREL	Cover design
Eileen Ferrance, NE&I	Activity 2.2
Joan Herman, CRESST	Chapter 2
Wendy McColskey, SERVE	Chapter 6; Activity 6.1; Sample 6.1; editing
Ann Muench, FWL	Activities 4.2, 5.4; Sample 3.1; editing
Jackie Palmer, SEDL	Chapters 3, 7; Activities 2.2, 2.9, 3.2; editing
Kit Peixotto, NWREL	Chapter 7
Kathy Petersen, NWREL	Word processing, editing, production
Mahna Schwager, SWRL	Editing

Table of Contents

The authors have formatted the *Toolkit* for ease of use in professional development. Thus, each chapter, activity, and appendix has been numbered separately to facilitate reproduction for handouts. Listed here are simply the chapter, activity, and appendix titles.

To assist users in finding information, a detailed table of contents is included at the beginning of each chapter. Additionally, there are detailed footers on each page to facilitate re-assembly.

Chapter 1: Introducing the *Toolkit*

Chapter 2: Supporting Educational Improvement with Alternative Assessment

Activity 2.1: *Setting the Stage for Becoming a Good Consumer of Science/Mathematics Alternative Assessments*
Activity 2.2: *Pictionary*
Activity 2.3: *Post-it Notes™*
Activity 2.4: *Seeing Wholes*
Activity 2.5: *Clapping Hands*
Activity 2.6: *A Comparison of Multiple-Choice and Alternative Assessment*
Activity 2.7: *Think of a Time*
Activity 2.8: *Sam's Story*
Activity 2.9: *Going to School*
Activity 2.10: *Tasks vs. Performance Criteria*

Chapter 3: Integrating Assessment with Curriculum and Instruction

Activity 3.1: *Camping Trip: A Math Problem-Solving Task*
Activity 3.2: *Sorting Fish: A Science Task*
Activity 3.3: *Miss Tolivar's Math Class*

Chapter 4: Exploring Design Options for Alternative Assessment

Activity 4.1: *Performance Tasks—Keys to Success*
Activity 4.2: *Spectrum of Assessment Tasks*
Activity 4.3: *Performance Criteria—Keys to Success*

Chapter 5: Evaluating the Quality and Equity of Alternative Assessments

Activity 5.1: *How to Be a Good Consumer of Science/Mathematics Alternative Assessments - or - How to Tell the Gems from the Duds*
Activity 5.2: *Dr. Know-It-All*

Chapter 5—continued

Activity 5.3: *Chickens and Pigs: Language and Assessment*
Activity 5.4: *Equity Role Play*
Activity 5.5: *Questions About Culture and Assessment*

Chapter 6: Using Alternative Assessment in Grading and Reporting

Activity 6.1: *Weighty Questions*
Activity 6.2: *How Can We Know They're Learning?*

Chapter 7: Designing Effective Professional Development

Appendix A: Alternative Assessment Sampler

Appendix B: Resources

Chapter 1

Introducing the *Toolkit*

Rationale for and Purpose of the *Toolkit*

In response to publicized reports that U.S. students were lagging behind in international comparisons of academic performance, along with projections of changing knowledge and skills needed to compete in a global workforce, numerous efforts at reforming mathematics and science education have arisen. (Appendix B provides a brief list of reform efforts and other pertinent information for further reading.)

While the list in Appendix B is by no means exhaustive (it should definitely be supplemented by local site-based management efforts), it does indicate the scope of the reform effort in mathematics and science education. The reform effort is fueled by research in education and psychology which supports a changing view of how learning occurs. In this new perspective, the learner actively constructs personal meaning from information and experiences by linking new information with his or her pre-existing knowledge and understanding. This "constructive" learning process requires changes in the interactions between teachers and students in the classroom—for example, more self-reflection, group collaboration, and teacher as facilitator. This gives students more opportunities to rearrange, add to, or delete information from their "personal conceptual frameworks."

Since assessment is intimately linked with instruction, it too is changing. Alternative forms of assessment are being used to assess different kinds of student outcomes (e.g., critical thinking and problem solving), to help teachers deepen their visions of what students need to know and be able to do, and to change the interactions between teachers and students in the classroom. The *Toolkit* has been designed specifically as a professional development resource to assist in this effort. It is a compilation of activities and supportive materials developed not only as an alternative assessment resource, but as a means of engaging teachers in dialogue about changing the way mathematics and science have traditionally been taught and assessed. It is designed to stimulate conversation among educators about one aspect (assessment) of a much larger reform effort.

In this *Toolkit*, we use the following definition of *alternative assessment.*[*]

> *Alternative assessment includes any type of assessment in which students create a response to a question rather than choose a response from a given list (e.g., multiple-choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios.*

Specifically, the goals of this *Toolkit* are to:

1. Provide information on the basics of assessment
2. Provide background/foundation information on alternative assessment
3. Provide information on the types of assessments that are being developed around the country and world and illustrate these variations with actual samples
4. Discuss the issues and considerations surrounding the development and use of alternative assessments so that educators become good consumers and users of current instruments and good developers of new assessments
5. Emphasize not only the monitoring functions of alternative assessments, but also the instructional potential
6. Assist users to develop a vision of what they want their assessments to look like (and have reasons for these choices)
7. Provide professional development activities that allow educators to construct their own understandings of the nature and role of alternative assessment

Many of these topics are appropriate both for assessment that takes place in the classroom by individual teachers and for assessment that occurs in more standard ways across classrooms and throughout districts and states (large-scale assessment). We will note when a particular consideration or practice is more appropriate for large-scale or classroom assessment.

Even though, in this document, we are emphasizing alternative assessment, we do not want to imply that only alternative assessments are worthwhile and all previous assessment efforts are worthless. Actually, no single assessment approach can accomplish all goals for assessment. A balance must be built by carefully considering all the types of things to be assessed and the reasons to assess them, and then choosing the assessment approach that best matches targets and purposes. Sometimes the answer will be an alternative assessment, sometimes traditional assessment, and sometimes a combination of the two.

---

[*] An interesting historical note is that 60 years ago multiple-choice was considered an alternative to the current assessment of the time—subjective, teacher-based judgment. The attempt was to be more "objective" by reducing subjective judgment. Now, "alternative" is used to mean assessments other than multiple-choice, true/false, and matching. However, the issue of being "objective" is still the same as 60 years ago. The goal now is to make subjective assessment (performance and other alternative assessment) as objective as possible.

Content of the *Toolkit*

The *Toolkit* includes the following:

Chapter 2: Supporting Educational Improvement with Alternative Assessment. The goals of this chapter are awareness of what constitutes good assessment, building a common language for assessment, and foundation knowledge. The chapter includes a discussion of the rationale for alternative assessment in science and mathematics, definitions of terms, discussions of purposes for assessment and the need for clear student learning goals, and help with deciding when alternative assessment should be used.

Chapter 3: Integrating Assessment with Curriculum and Instruction. The goal of this chapter is the understanding of the various ways that development and use of assessment can affect and enhance instruction. It includes vision building on how performance assessments can be useful instructional tools if they are designed properly.

Chapter 4: Exploring Design Options for Alternative Assessment. This chapter provides a summary and analysis of current alternative assessment efforts. The concepts developed in this chapter are used throughout the remaining chapters. Actual assessment instruments are used to illustrate the points made.

Chapter 5: Evaluating the Quality and Equity of Alternative Assessments. The goals of this chapter are to understand the meaning and importance of quality, and to understand equity issues. Guidelines that delineate characteristics of sound alternative assessment are provided. These can be used when either selecting or developing alternative assessments. Samples from actual assessments are used to illustrate the points made.

Chapter 6: Using Alternative Assessment in Grading and Reporting. Grading and reporting are two issues that consistently recur in discussions of alternative assessment. This chapter discusses the accommodation of grading and ways to report assessment results that reinforce instructional uses.

Chapter 7: Designing Effective Professional Development. The goals of this chapter are to give an overview of how to provide effective professional development, and to supply sample professional development agendas and vignettes using portions of the *Toolkit* to illustrate professional development options. The chapter includes forms for evaluating professional development activities.

Appendix A: Alternative Assessment Sampler. Appendix A contains material from 21 different assessment projects.

Appendix B: Resources in Alternative Assessment. Appendix B provides references on additional resources in alternative assessment—reform projects, organizations developing assessments, print materials, and videotapes.

Each chapter has several parts:

- A chapter introduction that includes goals for the chapter, an outline of chapter content, and an index of the professional development activities, sample assessments, and other chapter resources included or referred to in the chapter
- A written section that presents information on the concepts and ideas in the chapter
- Associated professional development activities, if appropriate

Use of the *Toolkit*

As the name implies, a toolkit contains a number of different tools that can be used to construct a multitude of products. The activities contained in *Improving Science and Mathematics Education—A Toolkit for Professional Developers: Alternative Assessment* are like the hammers, saws, and screwdrivers of a carpenter's toolkit. When used creatively and with other necessary elements, they can lead to a useful and successful product. This *Toolkit* is not, however, intended to be a complete textbook on assessment. Nor is it intended to provide all the necessary background and expertise to mount effective assessment development. Rather, it provides assistance on some of the major areas in which educators have questions.

The *Toolkit* chapters are meant to be both stand-alone information pieces and resources for professional development. *As a professional development resource, the* Toolkit *is intended to complement, not substitute for, local assessment expertise.* Although we attempt to present highly technical topics in simple and accessible ways, facilitators and trainers will need technical expertise and experience to address the multitude of questions that often arise during professional development activities.

All materials needed for professional development, including hard copies of overheads, are included, as appropriate. The activities included in this *Toolkit* are designed to model effective teaching strategies incorporating what we know about good teaching practices in mathematics and science education. However, *your* purpose, setting, participants, time, and resource considerations will affect the actual planning and implementation of your professional development opportunity. So please feel free to modify exercises and activities to meet your own needs.

Who Developed the *Toolkit*?

The Regional Educational Laboratory Network Program

During the nearly three decades since their inception, the regional educational laboratories (funded by the Office of Educational Research and Improvement of the U.S. Department of Education) have proven to be valuable resources in their regions. Each laboratory identifies regional needs and develops programs to help meet them. In cooperation with partners in state and intermediate education agencies, universities, professional associations, foundations, business and other social service agencies, the laboratories provide programs and services to schools and others working to improve education.

In 1992, the Regional Educational Laboratory Network Program was established in recognition of the growing need for coordinated national responses to America's educational challenges and the potential of the laboratories working collaboratively to help meet this need. All 10 have joined together to formalize, consolidate, and extend their capability to act as a national system.

The structure for achieving this goal is a set of collaborative projects, staffed and supported by all or a subset of the regional laboratories. Each project has an originating (or "lead") laboratory which provides a project coordinator. The coordinator forms a steering committee (called the design team) to shape the project plan and activities. Collaborating laboratories then provide one or more staff to help carry out the project.

The content emphases of the projects are mathematics and science, communications development, system building, and underserved populations. Examples of current project topics are:

- Alternative assessment database
- Professional development toolkits
- Native American education promising practices
- Teaching cases professional development

- Multimedia school improvement resource system
- Urban and early childhood networks
- Sharing promising and proven practices

The Laboratory Network Program on Alternative Assessment (LNP-AA)

The *Toolkit* was developed by the Science/Mathematics Alternative Assessment Laboratory Network Program (LNP-AA). The goal of this project is to collect and make available information about alternative assessments by pooling the resources of the regional labs and CRESST (Center for Research on Evaluation, Standards and Student Testing at UCLA).

In addition to the *Toolkit,* the LNP-AA also has available the following products and services:

1. **Improving Science and Mathematics Education—A Database and Catalog of Alternative Assessments, 2nd Edition.** The database contains descriptive information on alternative assessment instruments, procedures, and ideas. Information in the database includes who to contact, what the assessment covers, grade levels, the types of tasks given to students, how responses are scored, and the availability of technical information. Many of the samples presented and described in the *Toolkit* were gathered as part of the database effort. The LNP-AA will update the database on a yearly basis through 1995. The database is available on disk (for either IBM or Macintosh using FileMaker Pro software) or over the Internet.

2. **Annotated Bibliographies.** Articles related to alternative assessment in science and mathematics, but not appropriate for the database, are collected in annotated bibliographies. Articles include such topics as the rationale for alternative assessments and what we should teach/assess. Bibliographies are available in print form and electronically over the Internet.

3. **Assessment Instruments.** Some regional laboratories have hard copies of the instruments in the database available for examination. (If you want to actually use one of the assessments you must contact the developer.)

Please contact your LNP-AA representative or Northwest Regional Educational Laboratory if you are interested in these additional products. (See the map and list of contacts at the front of the *Toolkit*.)

Chapter 2

Supporting Educational Improvement with Alternative Assessment

What's in this Chapter?

Assessment of student achievement in mathematics and science is changing because students face a world that will demand new mathematics and science abilities. Today's assessments need to measure not only the basics, but also a student's ability to think critically, analyze, and make inferences—skills found in the standards adopted recently by national commissions in mathematics and science. In addition, today's assessments need not only to measure but also to support new methods of instruction.

This chapter explains in more detail why changes in assessment are taking place and what they mean for schools engaged in the educational improvement process. The activities at the end of this chapter are designed to elicit participants' current familiarity with and level of understanding about assessment and to stimulate dialogue about alternatives that do not fit the structure and format of traditional assessment.

Chapter Goals

1. Establish background knowledge about assessment
2. Increase awareness of the principles of good assessment
3. Lay the groundwork for good assessment as a tool for educational improvement
4. Agree on a common language of assessment terms

Chapter Content

Readings

A Vision for the Role of Assessment in Educational Improvement .... Page 6

This section introduces the idea that good quality assessment information is necessary for educational improvement. As goals for students' performance change, so too does information needed for decisionmaking. We need quality assessment information that can (1) be used to make decisions about the need for and the effectiveness of improvement strategies, and (2) contribute to the design of those strategies.

Rationale for Change in Assessment .... Page 8

The rationale for changing assessment, as set out in the first section of the chapter, is described in more detail. Subsections deal with changes in valued student outcomes, changes in our notion of competence, changes in what we want assessment to communicate, changes in the purposes of assessment, changes in the importance of relevancy, and changes in the recognition of the critical role of teachers in assessment.

High-Quality Assessment Starts with Knowing Your Purpose and What Is to Be Assessed .... Page 13

The discussion of quality in this chapter starts with consideration of the importance of purpose in determining the form of assessment to be used. Different types of assessments are more appropriate for some purposes than others. A high-quality, valid assessment is one that provides accurate information for intended purposes.

Ensuring Sound Assessment Practice—The Rest of the Story .... Page 16

Although subsequent chapters in the *Toolkit* deal with many of the technical quality issues in detail, this section of Chapter 2 introduces the steps in the assessment design or selection process that result in quality assessments. The steps include defining achievement targets to be assessed; determining the purpose of the assessment; matching the assessment method to the targets; selecting or developing appropriate assessment tasks/items; specifying the criteria and standards for judging student performance; developing or selecting a reliable rating process; avoiding the pitfalls that can lead to mismeasurement; collecting evidence of reliability and validity; ensuring consequential validity; using test results to refine assessment and improve curriculum and instruction; and providing feedback to students, parents, and the community.

Summary .... Page 18

Glossary .... **Page 20**

The glossary includes definitions of the assessment terms used in the *Toolkit*.

References .... **Page 25**

Activities in this Chapter

Activity 2.1 *Setting the Stage for Becoming a Good Consumer of Science/Mathematics Alternative Assessments*

Discussion of the rationale for alternative assessment and the vision of what alternative assessment in science and mathematics should be like. Time: 75 minutes

Activity 2.2 *Pictionary*

Light-hearted review of assessment terminology. Time: 30 minutes

Activity 2.3 *Post-it Notes™*

Review of assessment terminology and mixer. Time: 20 minutes

Activity 2.4 *Seeing Wholes*

Emphasizes the connection between assessment, curriculum, and instruction. Time: 20 minutes

Activity 2.5 *Clapping Hands*

Participants play the part of assessors and assessees to explore the consequences of designing alternative assessments in various ways. Time: 75 minutes

Activity 2.6 *A Comparison of Multiple-Choice and Alternative Assessment*

Participants compare a multiple-choice test to an alternative assessment that attempts to measure the same skills, and discuss the advantages and disadvantages of each approach. Time: 60 minutes

Activity 2.7 *Think of a Time*

Participants reflect on their own experiences being tested and draw conclusions about what they want their assessments to be like. Time: 20 minutes

Activity 2.8 *Sam's Story: A Comprehensive Assessment*

Participants use a succession of assessment information about Sam to draw more and more sophisticated conclusions about his achievement. Time: 45 minutes

Activity 2.9 *Going to School*

Demonstrates the importance of performance criteria, illustrates different types of performance criteria, and discusses the advantages and disadvantages of various types of rubrics. Time: 60 to 90 minutes

Activity 2.10 *Tasks vs. Performance Criteria*

This discussion assists participants with looking at some examples of alternative assessments and illustrates the two main parts of an alternative assessment—tasks and performance criteria. Time: 20 minutes

Related Activities in Other Chapters

Activity 3.1 *Camping Trip: A Math Problem-Solving Task*

Demonstrates how to develop performance criteria and illustrates how the development of performance criteria can be instructional for both teachers and students. Time: 60 to 90 minutes

Activity 3.2 *Sorting Fish: A Science Task*

This activity is similar to Activity 3.1, but uses a science task instead of a math task. Time: 60 minutes

Activity 4.1 *Performance Tasks—Keys to Success*

Illustrates the dimensions along which performance tasks differ. Discusses the advantages and disadvantages of different approaches. Time: 60 to 90 minutes

Activity 4.2 *Spectrum of Assessment Tasks*

Participants review a variety of short, related assessment tasks and rank them according to amount and type of information each might elicit from students. Time: 60 minutes

Activity 4.3 *Performance Criteria—Keys to Success*

Illustrates the characteristics of sound performance criteria. Time: 60 minutes

Activity 5.1 *How to Be a Good Consumer of Science/Mathematics Alternative Assessments - or - How to Tell the Gems from the Duds*

This activity gives participants practice with making judgments about the quality of the science and mathematics alternative assessments they may want to try. Areas of examination include coverage, performance tasks, performance criteria, equity and fairness, consequences, and cost. Time: 6 hours

Activity 5.2 *Dr. Know-It-All*

This video-based activity illustrates with humor many subtle and not-so-subtle biases and messages sent by instruction and assessment activities in science. Discussion questions focus on how issues raised by the video may apply to the assessments participants are examining. Time: 30 minutes

Activity 5.4 *Equity Role Play*

Participants explore their personal beliefs, attitudes, and values by creating and playing the role of a student or teacher involved in potentially inequitable assessment activities. Time: 90 minutes

Activity 5.5 *Questions About Culture and Assessment*

This activity is designed to increase awareness of the relationship between culture and assessment and demonstrate how cultural factors can affect students' ability to show what they know and can do. Time: 90 minutes

Activity 6.2 *How Can We Know They're Learning?*

A version of Activities 3.1 and 3.2 designed for parents.
Time: 2 hours

Supporting Educational Improvement with Alternative Assessment—Readings

A Vision for the Role of Assessment in Educational Improvement

The story of education in the 1990's continues to be one of reform and change. Both our goals for students and the strategies we use to help students reach those goals are changing. We are moving from an industrial age, where knowing basic reading and arithmetic skills was all that was necessary for success, to an information age where accessing, interpreting, analyzing, and using appropriate information for decisionmaking is required. What is needed for success in life is also leading to changes in the goals we have for students. In the past, learned individuals knew a lot of facts and were well informed about the world around them. Today, with the volume of information doubling every three years, there are too many facts to learn. Rather, students will need to know how to access the information they need and how to apply it to increasingly ambiguous, real-life problems. Students will face many situations where there will be no clear-cut right answer, but where, instead, they will need to analyze the situation and apply their knowledge and skills to find a solution that will work. Yes, knowledge and facts may still be important, but they simply are not sufficient to prepare students for future success.

At the same time, research tells us that different instructional approaches are required if students are to learn the complex thinking and problem-solving skills they need. In the old system of education, teachers were viewed as providers of knowledge and students as passive recipients of that knowledge. Although such attitudes persist in some sectors of education, today we know that good instruction engages students actively in the learning process, requiring them to construct their own meanings and actively acquire the knowledge bases they use.

What do these changes have to do with assessment? Through assessment we communicate to students, teachers, and the public about what is important to teach and learn in school. The often heard query "What's going to be on the test?" is an important one. The answer tells teachers, no less than students, where to concentrate their efforts. But, contrary to current understandings of students' needs, many of our traditional multiple-choice and true-false tests have focused on the testing of facts and skills in isolation, seldom requiring students to apply what they know and can do in real-life situations. Over-reliance on these forms of assessment has led to an over-emphasis on basic knowledge and skills in instruction. Although these skills in themselves may be important goals of the educational process, other more complex skills have been ignored.

As our goals for students change and the demand for educational improvement intensifies, assessment tools must also change. We need assessments that are well matched with the full range of goals we hold for student performance. We need assessments to measure progress toward these goals and to provide the data educators need to make informed decisions about what and how to improve education. Good assessment data, well matched to goals and appropriately interpreted, better enable us to make good decisions.

Assessment, after all, is "the process of collecting, synthesizing and interpreting information to aid in decision making" (Airasian, 1991). It is an essential tool for anyone making decisions about educational improvement, from state policymakers trying to stimulate reform and assure accountability, to classroom teachers trying to assure the highest-quality instructional program for their students, to students and their parents who want to know "How am I doing?" and "How can I do better?" Whether we are state or local policymakers, school administrators, or teachers, assessments can serve a number of important functions that are essential in educational improvement. They help to establish and communicate standards, create instructional pathways, provide diagnostic feedback about students and programs, assess and evaluate the progress of students and programs, and communicate progress to others. And most importantly, when we have a sensible assessment system, assessment can support the improvement of education, by focusing all parts of the educational system on valued goals and providing feedback that enables everyone to make concerted efforts to achieve those goals.

This *Toolkit*, and the Laboratory Network Program that led to its creation, recognizes the important role assessment plays in the educational improvement process. Good quality assessment targeted at important goals for students supports planning for educational improvement. On the other hand, assessment that focuses on the wrong goals can impede reform.

For all these reasons, educators and educational policymakers are working to change the assessment of student achievement to reflect changes in the expectations and standards for students, in the methodologies and purposes of assessment, and in the instruction that assessment is intended to support. In this *Toolkit*, we refer to this evolving type of assessment as *alternative assessment*. As noted in Chapter 1, the definition we rely upon is:

> *Alternative assessment includes any type of assessment in which students create a response to a question rather than choose a response from a given list (e.g., multiple-choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios.*

Also, as noted in Chapter 1, even though this *Toolkit* emphasizes alternative assessment, this does not imply that only alternative assessments are worthwhile and all previous assessment efforts are worthless. The key to successful decisionmaking is to provide a comprehensive assessment of important student goals, using the most appropriate combination of assessment tools to do so.*

Rationale for Change in Assessment

As implied above, science and mathematics curriculum, instruction, and assessment are changing for a number of reasons. This section explains each of these reasons in more detail. They include changes in: valued student outcomes, our notion of competence, the authenticity or "real-worldness" of assessment tasks, the purposes of assessment, and recognition of the critical role of teachers in assessment. (Related activities are: **Activity 2.1**—*Setting the Stage for Becoming a Good Consumer of Science/Mathematics Alternative Assessments*, **Activity 2.4**—*Seeing Wholes*, and **Activity 2.6**—*A Comparison of Multiple-Choice and Alternative Assessment*

Changes in Valued Student Outcomes

Society's knowledge base has grown exponentially over the last few decades, and continues to do so. Scientists daily add new concepts and detail to our understanding of mathematical and scientific disciplines and their application in the medical, health, environmental, industrial, agricultural, technological and other areas.

> *Given this pace, no one individual can be expected to keep up with the information flow in a single discipline, much less across disciplines. Such a knowledge explosion makes futile most attempts to have students memorize and regurgitate large bodies of facts.*
>
> *Economic trends also push us away from a fact-based curriculum. The shift from a manufacturing- to an information- and service-based economy requires that individuals have skills in accessing and using information and in working with people. These changes in the workforce and in the pace and complexity of modern life suggest that people will need to be flexible, to shift jobs frequently, and to adapt to change. To prepare students for success in the future, schools must emphasize how to apply rather than just acquire information.* (Herman, Aschbacher & Winters, 1992, p.14)

Goals for students thus are changing. For example, the National Council of Teachers of Mathematics has released a set of national standards for mathematics (Commission on Standards for School Mathematics, 1989), and the National Research Council (1992) is drafting standards for science. Other groups are involved in the reform of mathematics and science education (for example, National Science Teachers' Association, 1992; AAAS, 1989; and Rutherford & Ahlgren, 1990). In these efforts, expectations for students reflect an increasing emphasis on critical thinking, problem solving, the ability to monitor one's own performance, the efficiency with which tasks are accomplished, group collaboration, and communication skills. These are difficult to measure in multiple-choice format, and practitioners need alternate ways to assess student competency on these outcomes.

---

* For this reason, multiple sources of assessment information about student achievement, including assessments that will allow for the evaluation of complex skills, are important. No single type of assessment will provide all the information that teachers need. In this new era of assessment, all of the assessment techniques and strategies that are available will be needed.

> *In a thinking curriculum, students develop an in-depth understanding of the essential concepts and processes for dealing with those concepts, similar to the approach taken by experts in tackling their tasks. For example, students use original sources to construct historical accounts; they design experiments to answer their questions about natural phenomena; they use mathematics to model real-world events and systems; and they write for real audiences.* (Herman et al., 1992, p.17)

In mathematics, for example, knowing how to add, subtract, multiply, and divide was sufficient for a simpler time when keeping a checkbook was the most advanced skill expected of individuals. Today, mathematics plays a larger role in the workplace and in everyday life. Knowing how to apply appropriate mathematical problem-solving strategies to real life situations (such as using computers to track and make sense of numerical data to solve a measurement problem at work, or to predict market demand for a product) is central to becoming a mathematically literate person.

Changes in Our Notion of Competence

Our notions of competency and expertise are also becoming more refined. For example, the definition of what it means to be a good reader has changed over time. Now it is not enough just to be able to convert symbols on the page to sound. To read at the levels necessary to be functionally literate in today's world, students need to be able to draw inferences, relate current information to past information, have a variety of efficient reading strategies, know when they are not understanding, and so forth. These skills are very difficult to measure in fixed-response format because it is not just a matter of testing separate skills, but also a matter of assessing a student's ability to know how and when to use a variety of skills collectively to accomplish a goal.

Similarly, traditional science courses required the learning of vocabulary and memorization of basic scientific principles. Today, we recognize that memorizing formulas does not necessarily result in "expertise" and may, in fact, be detrimental to the development of expertise. Students must understand and be able to apply those scientific principles to real-life problems. More science classrooms today look like laboratories where students are learning by doing. Science assessments are beginning to change from mere vocabulary tests to assessments that enable students to demonstrate their conceptual understanding of the material.

Changes in What We Want Assessment to Communicate

Teachers, parents, and students develop an understanding of what is valued in education by virtue of what is assessed and how it is assessed. For example, when multiple-choice tests are used exclusively, it gives the impression that there is only one right answer and that there is always a right answer. Assessments also communicate who has the right and the ability to judge quality. Kober (1993) notes the following messages from traditional science tests:

a. Because they have a single right or wrong answer, they reinforce the misleading conception of science as a static body of facts

b. Because their results rank individual student performance against that of a larger group, they perpetuate the notion that only a few students—the top scorers—are smart enough to pursue science

c. Because they sample a breadth of content in a superficial and unconnected way, traditional tests actually reward instruction that drills students on low level facts and vocabulary recognition

Current theories of instruction seek to change the power relationships in the classroom, with teacher as coach/facilitator rather than imparter of knowledge, and students as assessors of their own and each other's performances. Students learn best when they are actively involved in the learning process. We need to reinforce school practices that are based on current learning theories by making the practices a part of assessment tasks and scoring criteria.

> *Current evidence makes it clear that instruction emphasizing structured drill and practice on isolated facts and skills does students a major disservice. Insisting that students demonstrate a certain level of arithmetic mastery before being allowed to enroll in algebra or that they learn how to write a good paragraph before tackling an essay are examples of this discrete skills approach. Such learning out of context makes it more difficult to organize and remember the information being presented. Applying taught skills later when solving real-world problems also becomes more difficult. Students who have trouble mastering decontextualized 'basics' are often put in remedial classes or groups and are not given the opportunity to tackle complex and meaningful tasks.* (Herman et al; 1992, p.15)

Assessment must change if we are going to support these changes in instruction. What kind of reinforcement do teachers receive for making necessary changes in classroom practice if the outcomes assessed by the tests do not cover the full range of our desired outcomes?

Changes in the Purposes of Assessment

Assessment in the past was most frequently used for sorting: grading, selection into special programs, assignment to instructional groups, identification of the highest and lowest performing students, and so forth. Different skills were expected of students depending on their ability. Today, all students need to achieve high standards of performance to succeed in the world, and sorting can no longer be the primary goal of assessment.

Tyler (1986) suggests that there is a new view of student assessment, one that is built around new expectations of education in contemporary society. The old view assumed that educational opportunities had to be rationed; those who were college-bound and those who would be leaders needed to be afforded the best education possible, while other students would become effective citizens, helpful members of society, without much education. The industrial jobs of that time allowed those without education to provide for their families and lead satisfying lives. The role of educational evaluation in this economy was to sort students into those who needed more education and those who did not. That is no longer the case.

In the current world, assessment must help identify the assets of students on which effective educational programs can be built and to identify the characteristics that are likely to interfere with the student's learning so that the school or college may help overcome these difficulties. In today's information age, *all* students need to meet high standards in order to succeed in the world. Educational assessment can no longer primarily play a sorting role.

Today, knowing that one student ranks higher than another is less important than knowing how both students' performances compare with our ultimate goals for performance. It is important for assessments to describe student performance well enough so that students and teachers know how students are progressing toward agreed upon goals. The standard is no longer "doing better than the rest," but rather meeting a previously agreed upon, rigorous level of knowledge and skills.

Changes in the Importance of Authenticity

Multiple-choice test results are "stand-ins" or proxies for the actual performance of interest. For example, we could use a multiple-choice test to assess knowledge about how to carry out a science experiment. If the student does well on the knowledge test, we might infer that the student could actually *do* the experiment. Performance on the multiple-choice test is therefore a proxy for the actual performance. However, although students who do well on multiple-choice tests of scientific knowledge tend to do better in the laboratory setting as well, the correlation does not hold for all students or all occasions. To use a multiple-choice test as the exclusive measure of a student's understanding of scientific principles ignores the need to assess the application of that understanding in the laboratory or some other real-life setting that requires scientific reasoning.

Modern theories of teaching and learning have demonstrated that students know, understand, and retain more when they learn it in the context of real-life situations. They can also demonstrate the depth of that understanding when the task they are asked to perform mirrors a real-life situation. Educators recognize the need to give students more realistic tasks to do on assessments because they want to know if the students can apply their knowledge to more real-life situations, and because we don't know whether artificial situations really elicit the full range of what students are able to do.

Changes in the Recognition of the Critical Role of Teachers in Assessment

More attention is being paid to helping teachers prepare themselves for classroom assessment and classroom uses of alternative assessment for a number of reasons:

1. Assessment is a frequently occurring activity in the classroom yet teachers rarely are given the opportunity to develop the skills they need to assess well and to use assessment information to improve instruction. As we will see in the next chapter, alternative assessments are useful as instructional tools.

2. Teachers are at the heart of educational improvement. It is they who will administer and interpret the results of performance assessments. If teachers don't feel a need for this information, or see how it can be used to improve student learning, their resistance may result in the same contamination of assessment results that occur with more traditional large-scale assessments.

3. Alternative assessment is not magic. Teachers will need to understand principles of assessment in order to ensure that new types of assessments are not misused or under-utilized. We need to build in those features that will make such assessments useful in the classroom, and then help teachers understand how to use them to improve the learning of their students. If we don't, all the wonderful "authentic" information our alternative assessments supposedly deliver may not materialize.

4. Changes that matter occur in classrooms, in the day-to-day interactions between teachers and students. Properly designed alternative assessments can support changes in these interactions. Our knowledge of how students learn is

increasing. Our assessments should reflect our new knowledge about teaching and learning.

With the increasing demands placed on teachers, support given through time and professional development is critical to the success of any educational improvement effort. This *Toolkit* is designed to provide one resource to help ensure that teachers will receive the support they need to use assessment in the service of instructional improvement.

High-Quality Assessment Starts with Knowing Your Purpose and What Is to Be Assessed

Purpose

Assessment results are only useful to the extent that they yield accurate and reliable information for the purposes for which the data were collected. The quality of an assessment is always determined relative to the assessment purpose. An assessment that is good for one purpose—for example, providing detailed diagnosis of a student's strengths and weaknesses—is not necessarily best for other purposes—for example, determining what the strengths and weaknesses of the school's overall curriculum are or whether most students attained the school's grade-level goals for student performance. Thus, a first major issue assessors encounter has to do with deciding on their purpose for assessment.

Just consider the differences in purpose between large-scale and classroom uses of assessment information. Large-scale purposes for assessment include accountability, selection of students into programs, certification of competence, graduation/promotion, and school rewards/funding. Classroom assessment purposes include grading, planning instruction, tracking student progress, evaluating the effectiveness of instruction, and promoting student control of learning.

These differences in purpose have implications for assessment design and use. For example, in general, large-scale purposes require more rigorous evidence of technical quality than do classroom assessments, primarily because important decisions are likely to be based on them. In contrast, for classroom purposes, a teacher has lots of formal and informal evidence upon which to base decisions, and so the results of any single, faulty assessment are not likely to be given undue weight.

It is inappropriate to use large-scale, standardized multiple-choice assessments designed for accountability purposes as a tool for instructional planning or to determine whether or not a student will succeed in college. Similarly, an individualized classroom assessment which allows a great deal of flexibility in administration and scoring would not provide enough uniformity to be a fair tool for statewide school accountability. Matching the assessment to the purpose is a critical first step in any assessor's choice of assessment method.

As other examples of how purpose can affect assessment design, consider these:

1. A single multiple-choice or short answer multiplication test may be perfectly acceptable to determine whether or not third graders have learned their multiplication facts, but would not be appropriate for making a decision about the overall quality of the third grade mathematics program.

2. A short answer or multiple-choice assessment designed to measure student knowledge of specific scientific principles might be useful for partially determining a student's grade. Inferring that this assessment sufficiently measures the student's ability to perform scientific tasks requiring an understanding of these principles would require observing the student applying that scientific knowledge in a laboratory setting.

3. Some assessments are used mostly to gather information about students in order to make decisions about them—for example, grading or certifying competence. Other assessments are designed more to involve students in their own assessment and thus serve an instructional function. This distinction between assessment to "monitor" and assessment to "teach" has implications for assessment design. (Chapters 3 and 4 discuss this more fully.)

4. Herman et al. (1992) describe two basic monitoring purposes for assessment. Different kinds of assessment are most appropriate for each purpose:

    a. The first purpose is to determine whether or not students have acquired specific knowledge or specific skills. The assessment should focus on the products of student learning using selected answer tests (such as multiple-choice) and direct assessment of projects and student products.

    b. The second purpose is to diagnose student strengths and weaknesses and plan appropriate instruction. Because we are interested in understanding where the student is going wrong, we need to assess the process as well as the product. Interviews, documented observations, student learning logs and/or self-evaluations, behavioral checklists, and student think-alouds are useful in conjunction with multiple-choice tests.

Content

It is also essential to have a clear picture of the student skills, knowledge and abilities to be assessed. For example, Marzano (1994) discusses the types of skills that are best assessed using traditional assessment and those best assessed using an alternative assessment. He also discusses implications for alternative assessment design of different types of student outcomes. Marzano divides student outcomes into two types: process skills and content/declarative knowledge skills. Figure 2.1 shows examples of these types of student outcomes.

Figure 2.1
When to Use Alternative Assessment

Type of Skills	Multiple-Choice, Short Response	Performance Assessment
Process Skills: Simple	Long division Punctuation, grammar Decoding words	
Process Skills: Complex		Problem solving Writing Setting up an experiment Critical thinking Group cooperation Lifelong learning Dance
Content/Declarative Knowledge: Simple	Recall facts—e.g., dates, places, events	
Content/Declarative Knowledge: Complex		Concepts—e.g., democracy Generalizations—e.g., "power corrupts"

Marzano states that both process skills and content knowledge can be simple or complex. Simple process skills are short routines that are applied fairly consistently across situations, such as long division. Complex process skills are those that require many decisions and the integration of many simple process skills, such as writing or critical thinking. Likewise, simple content knowledge relates to things like recall of facts, while complex content knowledge relates to concepts and generalizations. Examples are shown in Figure 2.1 above.

To assess **complex procedural skills**, Marzano would set up a task that requires students to use the skills in question and develop performance criteria to measure different levels of successful performance.

To assess **complex declarative knowledge**:

1. Decide how students will apply this knowledge—comparing, classifying, evaluating cases, deducing consequences, constructing support, etc.
2. Decide how students will report the results of their application—writing, speaking, drawing, etc.
3. Develop performance criteria for three areas:
   a. Content/declarative knowledge—how well does the student know the content?
   b. Quality of the product—how well did the student present the work in writing, speaking, etc.
   c. Quality of the application—how well did the student execute the knowledge application process, e.g., if the student was finding examples of a concept, how well was this done?

Thus, determining purpose is an essential first step in the selection or design of any assessment strategy.

Ensuring Sound Assessment Practice—The Rest of the Story

Sound alternative assessment begins with definition of content (the student standards or goals to be assessed) and the purpose for the assessment, as seen above. The assessment design process includes other steps, presented below (based on Herman et. al, 1992, p. 8). The steps of determining purpose and content are repeated on this list, just to make it complete. Many of these steps are covered in more detail throughout the *Toolkit*. Other topics are beyond the scope of the *Toolkit*. (We have noted those that are not addressed in detail.)

1. Clearly define what it is you want to assess (the achievement target).
2. Clearly state the purpose for the assessment, and don't expect one assessment to meet other purposes for which it was not designed.
3. Match the assessment method to the achievement target(s) and purpose, as seen in the previous section.

4. Specify illustrative tasks that would require students to demonstrate these skills and accomplishments. Avoid tasks that may only be interesting activities for students, but may not yield evidence of a student's mastery of the desired outcomes.

   (Task characteristics are discussed in Chapter 4. Two activities in Chapter 4 relate to task design: **Activity 4.1**—*Performance Tasks—Keys to Success* and **Activity 4.2**—*Spectrum of Assessment Tasks*. **Activity 2.7**—*Think of a Time* provides a very personal way of considering assessment design issues. **Activity 2.10**—*Tasks vs. Performance Criteria* provides examples of performance tasks.)

5. Specify the criteria and standards of judging student performance on the task selected in step 4. Be as specific as possible, and provide samples of student work that exemplify each of the standards.

6. Develop a reliable rating process that would allow different raters at different points in time to obtain the same, or nearly the same, results, or, if used in the classroom by a single teacher, allows for each student to be assessed using the same criteria.

   (Criteria and development of scoring processes are dealt with extensively in Chapters 3 and 4. Related activities are: **Activity 2.9**—*Going to School*, **Activity 3.1**—*Camping Trip*, **Activity 3.2**—*Sorting Fish*, **Activity 4.3**—*Performance Criteria—Keys to Success*, **Activity 6.2**—*How Can We Know They're Learning?* and **Activity 2.10**—*Tasks vs. Performance Criteria*)

7. Avoid the pitfalls that can lead to mismeasurement of students (threats to reliability and validity). Assessors should ensure adequate sampling of the content domain, avoidance of bias/subjective scoring, reasonable uniformity of assessment administration, avoidance of extraneous factors (e.g., too much reading required on a mathematics or social studies test), suitable environment for assessment, and awareness of and compensation for temporary student factors (e.g., parents' recent divorce or illness).

   (Pitfalls are discussed in detail in Chapter 5 including **Activity 5.1**—*How to Be a Good Consumer of Science/Mathematics Alternative Assessments*. A gentle introduction to technical issues, especially as related to classroom assessment, is found in **Activity 2.5**—*Clapping Hands*. Equity issues are explored in **Activity 5.2**—*Dr. Know-It-All*, **Activity 5.3**—*Chickens and Pigs*, **Activity 5.4**—*Equity Role Play*, and **Activity 5.5**—*Questions About Culture and Assessment*.)

8. Collect evidence/data that shows that the assessment is reliable (yields consistent results) and valid (yields useful data for the decisions being made). On performance assessments, this might be demonstrated through inter-rater

agreement on scoring and evidence that students who perform well on the assessment also perform well on other related items or tasks. With multiple-choice assessments, correlations should demonstrate internal consistency (students perform equally well or poorly on all related items) and should show that performance on the test correlates with performance on similar skills presented differently. In the classroom, where teachers have multiple measures of each of their students' performances, the formality of the collection of the technical quality data is not necessary.

This topic is not covered in detail in the *Toolkit.* (See the reference list at the end of this chapter for more assistance.)

9. Ensure "consequential validity"—the assessment maximizes positive side effects and minimizes negative ones. For example, the assessment should give teachers and students the right messages about what is important to learn and to teach; it does not restrict the curriculum; it is a useful instructional tool; and the decisions made on the basis of the assessment results are appropriate.

   (Consequential validity and the messages our assessments send teachers, students, and parents are discussed in Chapters 3, 4, 5 and 6. Related activities are: **Activity 5.1**—*How to Be a Good Consumer,* **Activity 5.2**—*Dr. Know-It-All,* and **Activity 5.4**—*Equity Role Play.)*

10. Use test results to refine assessment and improve curriculum and instruction; provide feedback to students, parents, and the community.

The Need for Professional Development

Very few schools of education provide coursework that enables teachers to understand and implement high quality assessments in their classrooms, or to understand how to use external assessment results to improve instruction (Stiggins, 1991). Intensive, ongoing professional development, and inclusion of assessment knowledge and practice into preservice education, will be increasingly necessary. Chapter 7 in this *Toolkit* is devoted to the topic of designing high-quality professional development.

Summary

This chapter provides the reader with an overview of the reasons changes are being made in assessment, and with a beginning understanding of the issues assessors must grapple with as they use (or develop) assessments to make educational decisions about students and programs. The remaining chapters provide a more indepth discussion of these topics.

Assessment in science and mathematics is changing because education in science and mathematics is changing. Research tells us that what students need to know and to be able to do, and the way in which knowledge and skills need to be taught, is often quite different from what schools now teach. This is not an indictment against schools. The changes in our world are happening at such a rapid pace that few aspects of society have been able to keep up. But keep up we must, if our students are to be successful in the changing world of today.

Although we present a strong case for alternative assessment in this *Toolkit*, we neither say that all assessments need to be of this type nor reject the use of multiple-choice and other forms of selected-response tests. We do affirm that alternative assessments, when designed well, offer appealing ways to assess complex thinking and problem-solving skills and, because they are grounded in realistic problems, are potentially more motivating and reinforcing for students. However, while alternative assessments may tell us how well and deeply students can apply their knowledge, selected-response (e.g., multiple-choice tests) may be more efficient for determining how well students have acquired the basic facts and concepts. A balanced curriculum requires a balanced approach to assessment.

The need for high quality assessment information to make informed decisions about changes in students and programs will be critical to the success of the educational improvement effort. There is no single correct method for assessing students, but there are ways in which all forms of assessment can be used well, helping schools make good decisions and meet students' needs.

Changes in assessment can promote and support the changes in mathematics and science curriculum and instruction that are required, but only if teachers are given the support, adequate planning time, and professional development needed to make the changes.

Glossary

The field of alternative assessment has created some unique language, much of which is used differently by different people. Reaching some common agreement on the meaning of the terms related to alternative and traditional assessments will help those using this *Toolkit* to help teachers understand the important concepts behind the terms. Without this common agreement, we may each think we know what another person means, but we may be wrong.

This glossary offers definitions of some of the most common terms used in this *Toolkit* having to do with:

- Types of assessments (e.g., alternative assessment, performance assessment, selected-response assessment)
- Student learner goals (e.g., benchmark performance, performance criteria, standards)
- Scoring techniques (e.g., analytical trait scoring, holistic scoring, scoring rubrics)
- Technical quality descriptors (e.g., generalizability, reliability, validity)

While all of these definitions may not appear in this chapter, they appear in other parts of the *Toolkit*.

A number of the definitions noted below are adapted from Hart (1993) and McTighe (1993). (Note: Assessment specialists sometimes disagree on the details of some of these definitions. We present the following list so that the reader will know what the authors of the current Laboratory Network Program *Toolkit* mean by these terms.)

(There are three activities in the *Toolkit* that help participants become more familiar with assessment terms:

1. **Activity 2.1**—*Setting the Stage for Becoming a Good Consumer of Science/Mathematics Alternative Assessment:* Includes a section on definitions.
2. **Activity 2.2**—*Pictionary:* Participants are asked to draw a picture of what the assessment term means and have others guess the term.
3. **Activity 2.3**—*Post-it Notes™:* Participants are given either an assessment term or a definition and try to find the person with a match.)

Definitions

Alternative assessment: any type of assessment in which students create a response to a question, as opposed to assessments in which students choose a response from a given list, such as multiple-choice, true/false, or matching. Alternative assessments can include short answer questions, essays, performance assessments, oral presentations, demonstrations, exhibitions, and portfolios.

Analytical trait scoring: a performance is judged several times along several different important dimensions or traits of the performance. Use of a scoring rubric and anchor papers for each trait is common. An example might be the judging of student problem solving for understanding the problem, correct use of procedures and strategies, and the ability to communicate clearly what was done.

Anchor papers or benchmark performances: examples of performances that serve as a standard against which other papers or performances may be judged; often used as examples of performances at different points on a scoring rubric. In math problem solving, for example, anchor papers are selected from actual student works that are considered to exemplify the quality of a performance level of 1, 2, 3, and so forth. If used with analytical scoring, there may be anchor papers or benchmark performances for each trait being assessed. Frequently there are also anchors for each grade level assessed.

Assessment: the act of collecting information about individuals or groups of individuals in order to understand them better.

Authentic (assessment): assessment tasks that elicit demonstrations of knowledge and skills in ways that resemble "real life" as closely as possible, engage students in the activity, and reflect sound instructional practice.

Benchmark performance: See "anchor papers." (Note: Project 2061 is using "benchmarks" to describe the science curriculum standards being developed. This is a different usage of the term than is intended here.)

Context (of an alternative assessment): the surrounding circumstances within which the assessment is embedded. For example, problem solving can be assessed in the context of a specific subject (for example, mathematics) or in the context of a real-life laboratory problem requiring the use of mathematical, scientific, and communication skills. Or, science process skills can be assessed in the context of a large-scale, high-stakes assessment or a classroom grading context.

Criteria: See "performance criteria."

Criterion-referenced assessment: an assessment designed to reveal what a student knows, understands, or can do in relation to specific performance objectives. Criterion-referenced assessments are used to identify student strengths and weaknesses in terms of specific knowledge or skills which are the goals of the instructional program.

Dispositions: affective outcomes such as flexibility, perseverance, self-confidence and a positive attitude toward science and mathematics. Some new assessments attempt to measure these outcomes.

Evaluation: a judgment regarding the quality or worth of the assessment results. Evaluations are usually based on multiple sources of assessment information. For example, "The information we collected indicates that students are performing above expectations."

Extraneous interference (or error): things that might cause us to mismeasure students, for example, excessive reading on a mathematics test, or role-playing on a science assessment.

Generalized rubric: a rubric that can be used to score performance on a large number of related tasks. For example, to score problem-solving and communication skills on *any* math problem-solving problem.

Generalizability: the extent to which the performances sampled by a set of assessment items/tasks are representative of the broader domain being assessed. For example, can we generalize about a student's problem-solving ability in general from the performance of the student on a specific set of 10 problem-solving tasks?

Holistic scoring: a single, overall score is assigned to a performance.

Indicator: a more specific description of an outcome in terms of observable and assessable behaviors. An indicator specifies what a person who possesses the qualities articulated in an outcome understands or can do. For example, a student may demonstrate his or her understanding of problem solving by finding a solution to a mathematics problem. The solution is an indicator.

Norm-referenced assessment: an assessment designed to reveal how an individual student's performance or test result ranks or compares to that of an appropriate peer group.

On-demand assessment: assessment that takes place at a predetermined time and place. State tests, SATs, and most final exams are examples of on-demand assessments.

Open-ended tasks: the kind of performance required of students when they must generate a solution to a problem or perform a task when there is no single, right answer. An example is: "Below you see a bar graph without any labels. What might this be a graph of?"

Open-response tasks: the kind of performance required of students when they are required to generate an answer, rather than select it from among several possible answers, but there is still a single, correct response. An example is: "There are four pieces of wood, each measuring seven feet. If you used them as a fence around your square yard, how large an area would you create?"

Performance assessment: direct, systematic observation of actual student performances and rating those performances according to pre-established performance criteria.

Performance criteria: a description of the characteristics that define the basis on which the response to the task will be judged. Performance criteria may be holistic, analytical trait, general or specific. Performance criteria are expressed as a rubric or scoring guide. Anchor papers or benchmark performances may be used to identify each level of competency in the rubric or scoring guide.

Portfolio: a purposeful, integrated collection of student work showing effort, progress, or degree of proficiency.

Primary trait scoring: a scoring procedure by which products or performances are evaluated by limiting attention to a single criterion or a few selected criteria. These criteria are typically based upon the trait or traits that are most essential to a good performance. For example, if a student is asked to write to the Department of Energy urging the opening or closing of a nuclear power plant, the primary traits might be the ability to communicate persuasively and the correct application of scientific knowledge to back up one's position. Scorers would attend only to these two traits.

Reliability: an indication of the consistency of scores across evaluators, over time, or across different versions of the test. An assessment is considered reliable when the same answers receive the same score no matter when the assessment occurs or how or who does the scoring, or when students receive the same scores no matter which version of the test they took.

Rubric: an established and written-down set of criteria for scoring or rating students' performance on tests, portfolios, writing samples, or other performance tasks.

Scale: the range of scores possible on an individual item or task. Performance assessment items are typically scored on a 4- to 6-point scale, compared to a scale of 2 (right/wrong) on multiple-choice items.

Selected-response assessments: students select the correct response from among a set of responses offered by the developer of the assessment. Multiple-choice and matching tests are examples of selected-response assessments.

Standardized assessments: assessments that are administered and scored in exactly the same way for all students. Traditional standardized tests are typically mass-produced and machine-scored and are designed to measure skills and knowledge that are thought to be taught to all students in a fairly standardized way. Performance assessments can also be standardized if they are administered and scored in the same way for all students. Standardization is an important consideration if comparisons are to be made between scores of different individuals or groups.

Standards (content or curriculum): statements of what should be taught. For example, the NCTM curriculum standards.

Standards (performance): an established level of achievement, quality of performance, or degree of proficiency expected of students. Examples include a cut-off score on a multiple-choice test or an expected benchmark performance on a performance assessment.

Task (as in a "performance task"): a goal-directed assessment exercise. For example, a particular math problem to solve, a lab to do, or a paper to write.

Task-specific rubric/scoring: a scoring guide or rubric that can only be used with a single exercise or performance task. A new rubric is developed for each task.

Validity: an indication of how well an assessment actually measures what it is supposed to measure rather than extraneous features. For example, a valid assessment of mathematics problem solving would measure the student's ability to solve a problem and not the ability to read the problem.

References

Airasian, P. (1991). *Classroom assessment.* New York: McGraw-Hill.

American Association for the Advancement of Science. (1989). *Project 2061: Science for all Americans* (Publication No. 89-01S). Washington, DC: Author.

Commission on Standards for School Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Hart, D. (1993). *Authentic assessment: A handbook for educators.* Reading, MA: Addison-Wesley.

Herman, J., Aschbacher, P. S. & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Kober, N. (1993). *EdTalk: What we know about science teaching and learning.* Washington, DC: Council for Educational Development and Research.

McTighe, J. (1993). *Glossary of assessment terms.* Frederick, MD: Maryland Assessment Consortium, c/o Frederick County Public Schools.

Marzano, R. (in press). Understanding the complexities of setting performance standards. In R. Blum & J. Arter (Eds.) *Now showing: A resource guide for student performance assessment in the context of school restructuring.* Alexandria, VA: Association for Supervision and Curriculum Development.

National Research Council. (1992). *National science education standards: An enhanced sampler.* Washington, DC: National Academy of Sciences.

National Science Teachers Association. (1992). *Scope, sequence and coordination of secondary school science, volume 1: The content core, a guide for curriculum designers.* Washington, DC: National Science Teachers Association.

Rutherford, F. J. & Ahlgren, A. (1990). *Science for all Americans—science literacy.* New York: Oxford University Press, Inc.

Stiggins, R. J. (1991, March). Assessment literacy. *Phi Delta Kappan*, pp. 534-539.

Tyler, R. (1986). Changing concepts of educational evaluation. *International Journal of Educational Research, 10.*

ACTIVITIES

Activity 2.1

Setting the Stage for Becoming a Good Consumer of Science/Mathematics Alternative Assessments

Purposes:

1. To help participants understand why it is necessary to change science and mathematics assessment
2. To help participants begin to create a vision of what they want assessment in science and mathematics to be like

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A2.1,O1** to **A2.1,O5**
- Chapter content to use as a handout: all of Chapter 2
- Paper and marking pens for participants

Time Required:

90 minutes to 2 hours

Facilitator's Notes:

A. Opening (15-20 minutes)

1. Display **Overhead A2.1,O1**, the title of the session.
2. Welcome participants to the session. Introduce yourself. Thank the sponsoring agency and anyone else who needs to be thanked.

3. Go over the goals for the session. Use **Overhead A2.1,O2**. Briefly mention how this training package came about (see Chapter 1).

4. Go over the handouts (Chapter 2 of the *Toolkit*).

5. Describe the other work of the Network: database, bibliographies, lending functions, etc.

6. Ask participants to briefly introduce themselves and mention the reasons that they are present. What particular information do they need? What are they hoping to get from the session?

B. Definitions (20-30 minutes)

We go over definitions so that, for this session, we are all using vocabulary the same way. Also, terms in assessment are changing rapidly right now; sometimes we have new words replacing the ones we have used before, for example, "dispositions" for "the affective domain." To proceed, you can simply select the three to six definitions you feel are most problematic for a particular group and define them in lecture mode. Or, you could have participants jot down what they think of when they hear these terms and then compare definitions. Alternative "definitions" exercises/icebreakers are **Activities 2.2** and **2.3**.

C. Vision Setting: Why Assessment Is Changing Right Now (45-60 minutes)

Use **Overhead A2.1,O3** and the handout in Chapter 2. The following might be one way to present this information. (Note: **Activity 2.6** also prompts discussion of why assessment is changing.)

1. Changes in Valued Student Outcomes

   We want to capture a richer array of what students know and can do. Ask the workshop participants to brainstorm a list of goals they have for students. The brainstormed list usually includes: critical thinking, lifelong learners, group workers, problem solvers, knowledge, facts, concepts, decision makers, good self-concept, motivation to learn, responsible, good citizens, metacognition, and flexibility.

   Summarize with **Overhead A2.1,O4**.

   Point out that this list is very consistent across the country right now. Ask participants which of these goals can be assessed in multiple-choice format and which can't. (Ask how targets would be measured differently by different approaches. Specifically, multiple choice can assess thinking skills, but does so "atomistically.") If we value these outcomes, and we list them in our district goals for students, how will we know when we have succeeded unless we assess differently?

2. Changes in Our Notion of Competence

   It is not just that our goals for students are changing. Additionally, the level of skill we expect from students is changing. Literate no longer means just being able to sign your name and read the newspaper. Students need to be able to read and process complicated documents, and use the information in thoughtful and creative ways. For all these reasons, we need more sophisticated measures of student outcomes.

   Ask the participants to return to their small group and select one of the goals for students. Ask them to describe a student they can remember who exemplifies the accomplishment of this goal. Have them list the student's attributes and accomplishments. Return to the large group and have several of the small groups report on their "standards." Point out those things that are assessable and explain how these can be used to help design an appropriate assessment.

3. Changes in What We Want Assessment to Communicate

   Our assessments communicate what we value. If we tell students that critical thinking is important, but our tests only have multiple-choice questions on facts, they will figure out pretty fast what is important to concentrate on. Likewise, if we tell teachers that science process skills are important, but our tests emphasize application of formulas, what will they concentrate on? Emphasize this point with the quote on **Overhead A2.1,O5**.

   Have participants brainstorm what the tests they use communicate to students, parents, etc. For example, consider standardized, multiple-choice tests that emphasize facts, curriculum-embedded tests that require students to produce responses, etc. Have them focus on more than just content; also, have them think about messages their assessments give students concerning how the world works (e.g., there is always one right answer to every question), and where power lies (e.g., only teachers have the knowledge and ability to judge progress).

4. Changes in the Purposes of Assessment

   Describe the differences in assessment purposes at the turn of the century (when only a few skilled workers were needed) and today (when all students need high levels of skills to succeed in the world). *Assessment in the past was most frequently used for sorting: grading, selection into special programs, rank order using norms, assignment to instructional groups, etc. Today, sorting is not the goal of assessment. Rather, we need to describe performance so that* *we know how students are progressing toward standards. Many of our current tests are not designed to do this.* Allow a brief time for questions and comments.

5. Changes in the Importance of Authenticity

   We want to give students more realistic tasks. First, our goal is to have students be able to apply what they are learning to their daily lives. How will we know if they can do that if we don't ask them to do it? Also, if the tasks are really artificial, we don't know whether they really elicit true performance. For example, will a timed writing sample with an assigned topic really show how well a student can write; or, do we need self-motivated pieces on topics of the student's choice?

   Have participants select the same student performance goal they discussed in #2 above. Ask them to think about what a student might be asked to do that would demonstrate his or her achievement of that standard. Share with the larger group and use this opportunity to point out the qualities of a good task.

6. Changes in the Recognition of the Critical Role of Teachers in Assessment

   We want to have better instructional tools and help teachers use them. This topic really leads into the next chapter—integrating assessment and instruction.

   Have participants discuss the usefulness of standardized tests for day-to-day teaching, e.g., how large-scale assessments offer policymakers information, but provide little useful information for teachers and, therefore, have little impact on the classroom other than to disrupt it with test preparation activities. What makes them less than useful? Would the situation change if only the type of assessment changed and not the way the assessment is developed or given (e.g., top-down)?

   It is our contention that large-scale assessments not only should *be designed with classroom usefulness in mind, they* must *be designed this way or we will run into the same problems as before. Currently, many assessments (norm-referenced and alternative) are designed for large-scale efficiency and not for classroom usefulness. A balance between these two purposes for assessment—measuring progress of students and schools and informing instruction—must be found. In Chapters 3 and 4 we describe those features of alternative assessments that make them more useful instructionally. Right now we only want to build the case that we should and must design large-scale assessments to be useful in the classroom, and we should and must provide teachers with the professional development they need to develop and use all of the assessment strategies at their disposal to get the best possible information about their students.*

a. Reasons we *should* design alternative assessments with instructional uses in mind: *First, alternative assessments, when designed properly, can be wonderful instructional tools, as will be seen in the next part of the workshop (Chapter 4). Second, scoring performances is a wonderful inservice mechanism (if the performance assessment is designed properly). Why not maximize benefit from all the time and money invested in a performance assessment?* Finally, ask the group, *Where does educational change occur? That's right, in classrooms. Assessments need to be designed in ways that promote change in the daily interactions between teachers and students. A properly designed performance assessment, used as an instructional tool, can change the interactions between teachers and students.*

b. Reasons we *must* design alternative assessments with instructional applications in mind: Ask the audience, *Who is it that administers our large-scale assessments? Teachers, that's who. What happens if they don't see the value or rationale in what they are doing? It might be that our wonderful, "authentic" performance assessments will not actually result in better information after all. If large-scale assessment developers don't take the needs of teachers into account, they might also not get what they want: better measures of student achievement for educational decisionmaking from the classroom to the state house.*

D. Closing

Summarize the purpose and goals of today's workshop with something like this: *This workshop was designed to help participants understand why changes in curriculum, instruction, and assessment are taking place, and how assessment is changing to promote these changes. Alternative assessment is designed to improve student learning of important mathematical and scientific principles heretofore reserved only for top achievers, to improve the information teachers have at their disposal for helping students and improving instruction, and to improve the match between what teachers teach and what is assessed so that teachers don't feel pulled in two directions (new curriculum and instruction, old assessments) and can focus on what's important: improving the education of children.*

Alternative assessment cannot occur in a vacuum. Policymakers and educators must accept and agree that changes in curriculum, instruction, and assessment are needed. To ensure the quality of these changes, professional development will be a key. This series of workshops on alternative assessment is intended to support the important work that is occurring in schools. We hope that you have found it helpful.

Setting the Stage: Why Changes in Science and Mathematics Assessment Are Needed

A Professional Development Session of the National Educational Laboratories and the Center for Research in Evaluation, Standards and Student Testing

A2.1,O1

LABORATORY
NETWORK PROGRAM

Goals for the Session

1. To understand why it is necessary to change science and mathematics assessment
2. To create a vision of what assessment in science and mathematics should be like

A2.1,O2

LABORATORY
NETWORK PROGRAM

Why Is Science and Mathematics Assessment Changing?

Changes in valued student outcomes

Changes in our notion of competence

Changes in what we want assessment to communicate

Changes in the purposes of assessment

Changes in the importance of authenticity

Changes in the recognition of the critical role of teachers in assessment

A2.1,O3

LABORATORY
NETWORK PROGRAM

> The definition of literacy that sufficed for an earlier age and a different economy has been replaced by a host of higher literacies...(that) call for students to analyze, think critically, evaluate, synthesize information, communicate more effectively, solve problems, learn how to learn, and, in general, learn far more actively than traditionally.

Rexford Brown
Testing and Thoughtfulness
Educational Leadership
April 1989

A2.1,O4

LABORATORY
NETWORK PROGRAM

What we choose to emphasize in our assigning of points is our de facto standard. It signifies what we really value, irrespective of what we profess to value....Is accurate but thoughtless recall, mere familiarity, worth more points than evidence of thoughtful understanding flawed by minor errors? How essential is the "form" of student work as opposed to its intellectual "content"? Is a particular student error important or merely easily noticed and counted?....In my experience, these questions are often unasked as teachers develop their tests.

Grant Wiggins
Rational Numbers
AFT, Winter 1988

A2.1,O5

LABORATORY
NETWORK PROGRAM

Activity 2.2

Pictionary

Purposes:

1. To get participants actively involved in the assessment workshop very early on
2. To introduce participants to each other and getting them to work together
3. To increase knowledge of assessment terms

Uses:

As an alternate to **Activities 2.1** and **2.3**

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overhead A2.2,O1**
- Paper, pencils, masking tape, colored marking pens, scissors, construction paper, glue
- Chapter content to use as a handout: Definitions

Time Required:

30 minutes

Facilitator's Notes:

1. Select several of the words or phrases from **Overhead A2.2,O1** or from the definitions list in Chapter 2 and display them on an overhead projector one at a time. The words on **A2.2,O1** are good for an initial awareness session.
2. Participants define the word/phrase by means of a drawing. They must not use any letters or numbers in the drawing. Allow about 30 seconds drawing time for each word.

3. After doing about five drawings, have participants share their sketch with others nearby. These will demonstrate how people visualize the meaning of a word/phrase and how people have different concepts of the same thing.

Alternative Option 1:

If there are participants with severe phobias against drawing in public, or people preferring another medium besides drawing, use the additional materials listed above.

Alternative Option 2:

Participants might not be limited to defining a word/phrase; they could also convey the concept represented by the word/phrase.

Alternative Option 3:

Divide participants into groups of three to five and give each group a slip of paper containing one of the assessment terms. Allow 5-10 minutes for groups to discuss and sketch the term/concept. Then have groups post their drawings on the walls. Allow a few minutes for participants to try to figure out the term/concepts being depicted. Then invite groups to explain their drawings to the large group.

performance assessment

performance criteria

performance task

authentic

standards (performance)

standards (curriculum)

A2.2,O1

LABORATORY
NETWORK PROGRAM

Activity 2.3

Post-it™ Notes

Purposes:

1. To expand or review understanding of assessment language
2. To develop a common language on which later activities can build

Uses:

In Chapter 2 as a mixer and to replace the definitional portion of **Activity 2.1**—*Setting the Stage*

Rationale:

It's important to acknowledge the realities of the evolving vocabulary for assessment—and to point out the importance of crafting shared meaning in order to move forward.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Small pieces of paper or Post-it™ notes containing definitions and terms (one possible set is included at the end of this activity, **A2.3,H1**)
- Chart paper and marking pens (useful, but not necessary)
- Chapter content to use as a handout: Definitions

Time Required:

10-15 minutes

Facilitator's Notes:

1. In this warm-up activity, participants examine assessment-related definitions and key terms. Each participant receives either a definition or a term. Their task is to find the partner who has the matching definition/term.

   The definition slips can be handed out individually as an opportunity for the facilitator to greet participants before the formal opening of the session, or can be included in individual packets of session materials. Participants have up to five minutes to find the partner in the group whose slip matches theirs. (If available, you can use one color for terms, another for definitions to facilitate the matching.)

2. Once participants have connected with the individuals they believe are their partners, ask each pair to place their term and definition on a chart until all terms and definitions are on the chart. With a large group, you can divide into smaller groups and have a chart for each group. With a relatively small group, review the intended matches, and clarify, solicit challenges, and refine meaning. With larger groups, you can have a walk-about in which participants seek confirmation of their choices in the work of other groups.

3. Close this brief activity with a summary of terms and, if appropriate, offer the opportunity for participants to write questions that still remain on a chart which will remain posted throughout the session. These questions often prompt significant thinking and give the facilitator some guidance about how to stretch following activities to take in the concerns of participants.

Using oral questions and student responses as opportunities for assessment; also includes interviews, examining student questions, and informal conversations.	Personal Communication
A question or task in which students must choose what they believe is the best or correct answer.	Selected-Response
A question or task in which students must create an answer or response themselves, but there is a specific correct answer or set of information that is expected.	Open-Response
Scoring performance as a whole; based on criteria, but not giving feedback on specific qualities of the student's work; giving a single score.	Holistic Scoring
A guide for scoring student performance; rules for scoring; sometimes includes descriptions of key characteristics of varying levels of performance.	Rubric

A2.3,H1

Scoring performance on several different dimensions; giving more than one score.	Analytical Trait Scoring
Data-based, systemic communication about student learning in writing, through portfolios, in three-way conferences, with numbers/letters or descriptive phrases, and using narratives.	Reporting
A question or task that has many correct or excellent responses—intended to promote thinking and application of ideas and information. There is no single, specific correct answer.	Open-Ended
Reporting a data-based summary of student learning to the public, including identification of systemwide strengths and weaknesses; and specific plans for using the information to improve learning.	Accountability
Systematically gathering information about student learning and using that information to communicate with students, parents, and others in order to improve learning and performance.	Assessment

A2.3,H1

Making judgments about student performance based on quality information gathered systematically over time.	Evaluation
The key qualities or dimensions of an effective performance.	Criteria
Short- or long-term activities that include rich opportunities to learn and systematic opportunities to assess the quality of student work.	Performance Tasks
Student thinking about their learning; recording or communicating about the process and content of learning.	Self-Reflection
Assessment that is based on observation and judgment of student created products and/or performances; intended to provide a rich portrait of student learning.	Performance Assessment

A2.3,H1

Activity 2.4

Seeing Wholes[*]

Purpose:

To engage participants in an examination of the connections among assessment, instruction and curriculum

Uses:

1. As an opener (or brief reminder) in the early stages of a multi-activity professional development session

2. Used with other activities (e.g., **Activity 2.3**—*Post-it™ Notes* and **Activity 2.6**—*A Comparison of Multiple Choice and Alternative Assessment)* to construct understanding about, and shared language for, alternative assessment

Rationale:

Making clear connections between instruction and assessment is an essential part of the development and use of alternative assessments that will be used for ongoing communication with students and parents. In this activity, participants bring their own perspectives to a definition that could equally be attached to assessment, instruction **or** curriculum. This activity also tends to surface assessment issues.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overhead A2.4,O1**—*A Round of Jeopardy*

Time Required:

10-20 minutes

---

[*] Developed by Dr. Joan Bycoff Baron at PACE, Harvard University, and used with permission.

Facilitator's Notes:

1. Show **Overhead A2.4,O1**—*A Round of Jeopardy* and have participants ask a question for which the statement is an answer. Write responses on an overhead.

   Here's an example of responses generated recently:

   - What is life?
   - What is a performance task?
   - What is curriculum?
   - What is ongoing assessment?
   - What is a system of coherent assessments?
     assessment?
   - What is a portfolio?
   - What is a good unit?
   - What is peer review?
   - What is education?
   - What is performance

2. Points made by Dr. Baron to link the responses back to the concept of integrated curriculum and assessment:

   A good assessment system is your curriculum!

   A good set of assessment tasks has these features: they are coordinated, there is a sequence to them, kids learn the criteria, they become the owners of what quality work is, the act of learning the criteria is the curriculum. Learning events are the midwifery.

   Good curriculum **is** good assessment if you specify the criteria.

A Round of Jeopardy

Answer:

A coordinated sequence of engaging and important learning events through which all children come to understand what high quality work looks like and how to produce it.

What is it?

A2.4,O1

LABORATORY
NETWORK PROGRAM

Activity 2.5

Clapping Hands*

Purposes:

1. To look at different ways performance assessments can be designed
2. To explore the way that these designs affect how the assessor and assessee feel—the consequences of how we do assessment
3. To develop an awareness and a better understanding of the pros and cons of involving students in their own assessment

Uses:

1. In Chapters 2 and 5 to illustrate features of quality assessment
2. In Chapter 4 to illustrate design options and the effects these can have on teachers and students

Rationale:

Assessment is a deeply affective thing. It is about success and failure, about deeply held feelings, and about our sense of self-esteem. This activity is designed to show the impact of assessment both on performance and on the feelings at the receiving and rating ends of the process. It provides a graphic demonstration of the unintended consequences of how we design our performance assessments.

---

[*] With thanks to Ruth Sutton for a demonstration of the activity and approval to use it in the *Toolkit*. Contact Ms. Sutton at 29 Nevile Court, Nevile Road, Salford, M7 OPS, England, UK.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Eleven chairs set up in a semi-circle at the front of the room, five on either side of a central chair in which the facilitator sits
- Tables and chairs facing the semi-circle for other participants
- Pencil and pad of paper for the facilitator
- **Overheads A2.5,O1**—*Clapping Exercise* and **A2.5,O2**—*Clapping Institute Criteria*
- **Handout A2.5,H1**—*Performance Assessment—The Meaning of Quality*

Time Required:

60-90 minutes

Facilitator's Notes:

The activity is experiential—it asks participants to commit and invest themselves in the process. As such, it might make some participants uncomfortable. Therefore, it is extremely essential to manage the activity well. If working with people who know each other well, it needs careful management from the center because previous relationships and sub-agendas might get out of hand. This is usually not the case in groups of people who do not know each other.

Once this exercise has been done with a group, it can't be done again with the same group. The exercise requires at least 15 people.

A. Set-Up

1. Prepare the "certificate" template mentioned on p. 4, step 4.
2. Introduce the exercise by going over the goals using **Overhead A2.5,O1**—*Clapping Exercise*. Also, tell participants that five people will be asked to do a simple performance task, and five other people will assess the performances. At the end everyone will be asked to describe how it felt to be in the position they took. Ask for 10 volunteers to come up and sit in the 10 chairs arranged in a semi-circle. The rest of the participants will participate at

the end by questioning the volunteers on their feelings and adding their own observations.

3. Ask the volunteers to be alert to their feelings as the exercise progresses. The real point of the exercise is to explore the consequences of handling performance assessment in various ways. Therefore the thoughts and feelings of the volunteers are very important.

4. Record the names of the 10 volunteers in a seating plan so that everyone can address them by name. The panel on the left of the facilitator will be the assessors and the panel on the right will be the assessees.

B. Assessment Task Administration

1. Ask the last person on the right to clap:

   Rick, clap for us.

   Do not ask the assessors to rate the performance. Do not provide feedback of any kind—verbal or nonverbal. Just smile and say:

   Thank you.

2. Ask the second to last person on the right to clap:

   Judy, clap for us.

   Ask this person to leave the room and come back when asked. Assign another individual to get him or her when asked to do so.

   Ask the assessors to rate the performance.

   Assess Judy's clapping on a scale of 1-5, with 1 being low and 5 being high.

   Do not provide any assistance to the assessors on the criteria for assessment. Do not answer any questions about how it should proceed.

   Poll the assessors on their scores, record them, and compute an average. Tell the panel the average.

   Call the person back in. Do not provide any feedback to him or her on what went on during his or her absence. Especially, do not provide feedback on the score average.

3. Ask the next person to clap:

   Bobbie, clap for us.

Send this person out of the room as before. Ask the assessors to rate performance on a scale of 1-5 with no discussion, as before. Compute the average. Call the person back in and tell him or her the score average:

OK, Bobbie, you got a ____.

Provide no other feedback.

4. Ask the next person to clap:

Doris, clap for us.

As before, send this person out of the room.

This time, the assessors get more guidance:

I've been in touch with the National Clapping Institute. They have sent us some criteria for judging clapping. (Show **Overhead A2.5,O2**—*Clapping Institute Criteria.*) *Unfortunately, the full handbook is not yet ready, so no additional detail is available.*

Assessors rate the performance on all three traits on a scale of 1-5 with no additional discussion or clarification of criteria. Resist the temptation to answer the questions of the assessors. Record and average the scores on each trait. Prepare a score report for the assessee:

Clapping Institute Award Certificate for Doris, (date)
Volume: (average score)
Appropriateness: (average score)
Creativity: (average score)

Bring the assessee back into the room and present the certificate with a handshake. Give no other comment except:

Here are your scores. I hope you'll find them useful.

5. Turn to the final assessee, and begin a dialogue about clapping. Ask such questions as:

Tell me about your previous experience with clapping. What kinds of situations do you find yourself clapping in?

You've had some experience of clapping (repeat what they said in response to the previous question). *What do you feel are your strengths as a clapper? Is there anything you feel you could change or would want some help with?*

Is there any guidance that you'd like to give the assessor panel—anything you'd like the panel to give feedback about?

Now turn to the assessor panel and begin a dialogue with them.

You're the experts on clapping, that's why you're here. Is there anything you want Ruth to know that you'll be looking at when she's clapping? Any discussion of the criteria given to us by the Clapping Institute? Anything you know is really important if she's going to be effective?

Now set a context for the assessment. Ask the assessee to describe the setting in which he or she would use the clap to be demonstrated in the upcoming performance and what he or she would be trying to accomplish with the clap. What kind of authentic atmosphere will he or she be clapping in? Then ask for any final questions or comments on the part of the assessors or assessees. Then ask the person to clap.

Ruth, please clap for us.

Ask the assessee whether or not he or she would like to leave the room. Ask the assessee how he or she would like feedback: verbally or as ratings on the various traits. Provide feedback in the way requested.

C. Debriefing

This part of the activity is essential. Participants will now have the opportunity to review how they felt about being treated in whatever manner they were treated.

1. Explain that each person will have the chance to describe how he or she felt during this process and the thoughts and realizations he or she had. After each individual has had a chance to talk, the floor will be open to the audience to ask questions or provide observations.

    a. Begin with the first assessee. Ask the general question:

    What thoughts or feelings did you have?

    Typical responses: Apprehension because I was given no guidance; every time I asked for clarification I was told, "We'll get back to you." The facilitator gave nothing. I kept wanting to try again after the criteria came out because I was sure I could do better. But, toward the end I actually felt glad I went first because by then the pressure was on.

    Follow-up questions:

    How did you feel when we came down the line and other people got other kinds of preparation? Feedback? How did you feel about getting no preparation?

Typical responses: Wanted to draw attention back to me. Felt it was increasingly unfair. Got angry.

b. Of the second assessee, ask the general question: *What thoughts or feelings did you have?*

Typical comments: I had to guess at the criteria, so fell back on general knowledge. Personality, cultural background, and gender all might have an effect when things are unclear. Tried to outdo Rick. I felt I was sent out of the room because I had done something wrong.

Follow-up questions:

How did you feel when you were asked to go out of the room? How did you feel about getting no feedback?

Now ask the panel how they felt at this time.

How did you feel about rating the second person and not the first? How did you feel about having no criteria?

Typical responses: Unsure of criteria so afraid of giving a high score. Angry with not having any criteria so decided to give everyone a "5". Angry that they were asked to rate the second person and not the first. Didn't have any criteria so only able to rate the second performance in relationship to the first. Felt sorry for the assessees.

c. After getting the general response from the third assessee, you could ask:

How did you feel when you were sent out of the room? When you got your score? Was that type of feedback useful?

Typical responses: A low score reinforced my self-judgment that I am a klutz. Angry that I was given no rationale for my score and no opportunity to improve it. A number with no rationale or scale is not useful. A gratuitous "5" from the judge rebelling because she didn't know the criteria was not useful, it felt condescending. Would have my "folks" come in and complain to the principal.

Ask assessors:

Did it bother you that I asked you to give Bobbie her scores and didn't ask you to give Judy her scores? Did you compare Bobbie with Judy?

d. After getting the general response from the fourth assessee, you could ask:

What did you feel like when you came back in and got your certificate? Was the feedback useful? How did you feel about the length of the wait? Could you hear people laughing? Did you think they were laughing at you?

Typical responses: Embarrassed that I was being treated differently. Embarrassed that some of my scores were low. Having a mental discussion with myself complaining about the exercise. Focused primarily on the low scores. Wanted to argue with the assessors.

Now ask assessors:

What were your reactions when I gave you criteria categories with no further discussion?

Typical responses: Not sure what "appropriate" or "creative" means. Made up my own idea of what each meant. Wanted to rate other things besides these three. The three categories did not capture all the relevant dimensions of performance. Not all three categories would be appropriate to assess in all contexts. The categories helped some.

e. After getting the general response from the final assessee, you could ask:

What did you feel like when I got to you and asked you to perform? Did you like the extra information? Did you feel uncomfortable being treated differently?

Typical responses: The pressure was on. How could I be better than the previous ones. Knowing I would be scrutinized by my peers made me nervous. The previous people had no guidance and so could be excused for poor performance, I had no such excuse. Embarrassed that I was treated differently. Afraid that if I got better scores the other "students" would be angry with me because they were treated differently. I wanted to not do a good job so that I wouldn't stand out.

A question for the raters:

Was the extra discussion useful?

Typical responses: Since we knew more about the person, we held her to a higher standard. Even though I specified my criteria for success, I still felt she should go beyond them to get a high score. I didn't want the assessee in the room while we gave scores because she might challenge my judgment.

f. After the discussion dies down, ask if any of the assessees would like another chance to clap.

2. Ask participants to reflect on this activity (e.g., what thoughts or insights they had). (Optional: Discuss in small groups.) Could also ask whether anyone has ever been in the position of one of the activity participants.

3. Ask the group to summarize what they learned about performance assessment. Summarize comments into the following categories:

   a. Things to be careful of when designing **tasks:** e.g., be careful of public performances; treat all students the same; put the performance into a realistic context; be specific on instructions; have open-ended tasks; have tasks that are meaningful to students; consider how to handle diversity and differences (e.g., special education, cultures, gender, learning styles).

   b. Things to be careful of when designing **performance criteria**: e.g., clear definitions; elaboration on how to assign points; agreement on what the criteria should be; models/examples; rater buy-in.

   c. Training of raters: e.g., practice and discussion are essential; sample performances are essential; have the raters do the assessment themselves; need consistency across sites; have raters explain their ratings.

   d. Student preparation: e.g., share criteria; give students the opportunity to add to criteria; train students on what the criteria mean and how it looks when performance is good and poor; use models; do formative assessments with feedback; make the assessment purpose clear; make sure students have prerequisite skills; make sure students trust the raters; teach students to do self-assessment using the criteria.

   e. Sources of mismeasurement: e.g., bias in tasks, criteria, or administration; criteria that don't cover the right "stuff"; poor training of raters; poor student preparation; tasks that don't elicit the right performance; sampling inadequacy; personality; embarrassment with regard to being compared to others; changing criteria; scheduling the assessment at bad times; fatigue of raters or students; lack of student or rater motivation; lack of teacher or student buy-in; rater bias; developmentally inappropriate assessment.

   f. Reporting results: e.g., make sure the scores have meaning; treat students the same; be specific and descriptive; provide evidence; allow time for discussion; emphasize what students *can* do, not what they *can't* do.

   Handout A2.5,H1—*Performance Assessment—The Meaning of Quality*, can be given to participants to summarize this discussion.

Performance Assessment—
The Meaning of Quality

Things to Be Careful of When:

Designing Tasks: be careful of public performances; treat all students the same; put the performance into a realistic context; be specific on instructions; use open-ended tasks; use tasks that are meaningful to students; consider how to handle diversity and differences (e.g., special education, cultures, gender, learning styles).

Designing Performance Criteria: clear definitions; elaboration on how to assign points; agreement on what the criteria should be; several models/examples; rater buy-in.

Training Raters: need practice, discussion, and sample performances; have the raters do the assessment themselves; need consistency across sites; have raters explain their ratings.

Preparing Students: share criteria; give students the opportunity to add to criteria; train students on what the criteria mean and how it looks when performance is good and poor; use models; do formative assessments with feedback; make the assessment purpose clear; make sure students have prerequisite skills; make sure students trust the raters; teach students to do self-assessment using the criteria.

Reporting Results: make sure the scores have meaning; treat students the same; be specific and descriptive; provide evidence; allow time for discussion; emphasize what students *can* do, not what they *can't* do.

Avoid Sources of Mismeasurement:

Mismeasurement occurs when a student's *real* ability is masked by irrelevant features of the assessment. Some examples of sources of mismeasurement are: features of tasks, criteria or administration that might affect one group of students more than others (often called "biasing features"); criteria that don't cover the right "stuff"; poor training of raters; poor student preparation; tasks that don't elicit the right performance; sampling inadequacy; student personality; embarrassment with regard to being compared to others; changing criteria; scheduling the assessment at bad times; fatigue of raters or students; lack of student or rater motivation; lack of teacher or student buy-in; rater bias; developmental inappropriateness.

A2.5,H1

Clapping Exercise

Goals:

Illustrate and reflect on different approaches to performance assessment

Explore the feelings of both assessor and assessee

Develop an awareness of the implications of student involvement in assessment

A2.5,O1

LABORATORY
NETWORK PROGRAM

Clapping Institute Criteria

Performance Criteria:

Volume

Appropriateness

Creativity

A2.5,O2

LABORATORY
NETWORK PROGRAM

Activity 2.6

A Comparison of Multiple-Choice and Alternative Assessment

Purposes:

1. To demonstrate and discuss the rationale for alternative assessment
2. To compare and contrast a traditional multiple-choice assessment with a performance assessment

Uses:

1. An alternative to the lecture/discussion presentation of rationale for alternative assessment in **Activity 2.1**—*Setting the Stage*
2. In Chapters 4 or 5 to illustrate task types. This could supplement **Activity 4.1**—*Performance Tasks—Keys to Success*

Rationale:

Alternative assessments are not meant to totally replace multiple-choice tests. Rather, each has its own advantages and disadvantages, strengths and uses. This activity is intended to enable participants to construct their own interpretation of these issues. It is also a good way to begin a general discussion about assessment. Frequently this will stimulate the questions that participants have on their minds about assessment. It also often results in a discussion of the rationale for alternative assessment.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Flip chart paper and colored markers
- Attached simulated standardized test page (this is modeled after, but not taken from, any current norm-referenced test), or use a page from the grade five math applications subtest of your local norm-referenced test
- Oregon's math problem solving performance assessment (**Sample 4.2**)

- **Handouts A2.6,H1**—*Mathematical Applications Test,* **A2.6,H2**—*Comparison of Multiple Choice and Alternative Assessment # 1,* and **A2.6,H3**—*Comparison of Multiple Choice and Alternative Assessment #2*
- **Overhead A2.6,O1**—*Summary*
- Chapter 2 content to use as a handout: Rationale for Alternative Assessment

This exercise could also be done with other grade levels and subject matter areas. Take a page from a standardized achievement test for the grade level and subject matter of interest. Then find a performance assessment with the following characteristics:

- A task that is relatively quick to read, e.g., **Samples 4.4**—*Sow Bugs,* **4.9**—*Mapping the Blue Part,* **4.10**—*Writing in Chemistry,* **5.1**—*Assessment and Technology Videotape,* or **5.2**—*Assessment of Laboratory Skills*
- Generalized performance criteria that have detailed descriptions, e.g., **Samples 4.8**—*Science—New Directions in Assessment,* **5.1**—*Assessment and Technology Videotape,* or **5.2**—*Assessment of Laboratory Skills in High School Science*

Time Required:

30-90 minutes

Facilitator's Notes:

1. Give participants the multiple choice and alternative assessment examples. In small groups have them analyze how they are the same and how they are different, and their relative advantages and disadvantages. You could have groups make two columns on a sheet of flip chart paper. In one column, have them list advantages and disadvantages (strengths and weaknesses) of the multiple-choice format; in the other, list advantages and disadvantages (strengths and weaknesses) for the performance assessment format. Have groups post their sheets. (10-15 minutes)

2. Large group discussion on the same topics. You could summarize group comments on a transparency or flip chart page (see **Overhead A2.6,O1**—*Summary).* (10-45 minutes)

   Handout A2.6,H3 comes from a discussion with teachers, administrators, curriculum coordinators, and education assessors. **A2.6,H2** comes from a discussion of *Toolkit* developers. They are included to illustrate the types of considerations that arise in these discussions. These can be used to assist in summarizing results or can be used as handouts.

3. Summarize, as needed, using the more formal presentation in **Activity 2.1**—*Setting the Stage* or the information in Chapter 2. (0-20 minutes)

MATHEMATICAL APPLICATIONS TEST

1. John and Sue are going to the movies. If they spend $12.00 for tickets, $3.50 for popcorn and $2.50 for drinks, how much money will they have left from a $20.00 bill?

   a. $12.00

   b. $8.00

   c. $6.00

   d. $2.00

2. An advertisement states, "Of doctors surveyed, 90 percent rate no other brand of sleeping aid better than Baye's." Which of the following could you conclude from this statement?

   a. Baye's sleeping pills may or may not be better than other brands.

   b. Baye's sleeping pills are the best.

   c. Almost all the doctors in the country prefer Baye's sleeping pills.

   d. Baye's is probably worse than other brands because advertisements always lie.

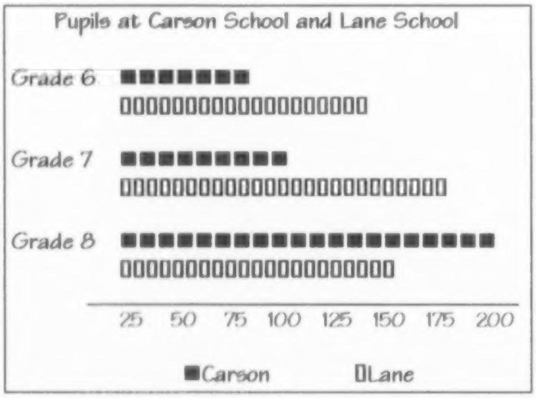
3. Quan Soo and Eddie have 9 baseball cards and 15 basketball cards. Eddie bought 6 more baseball cards. Which of the following statements is true?

   a. There are more baseball cards than basketball cards.

   b. There are more basketball cards than baseball cards.

   c. There is an equal number of basketball and baseball cards.

4. How many more eighth graders are there at Carson School than at Lane as shown on the graph below?



   a. 50 students

   b. 100 students

   c. 150 students

   d. 200 students

Comparison of Multiple Choice and Alternative Assessment #1

	Multiple Choice	Performance
Areas assessed		
• Individual process skills (e.g., interpreting graphs, addition, subtraction)	Yes	Yes
• Recall of facts	Yes	Can assess knowledge but not good for assessing the total range of knowledge for a course
• Sampling of skills and facts	Broad range	
• Understanding	Limited	Yes
• Higher-order thinking skills (e.g., critical thinking)	Limited	By design*
• Application of skills to solve real-world problems	Limited	By design*
Administrative considerations		
• Time to administer	Short to long	Usually long
• Cost-benefit	Inexpensive	More expensive
• Time and ease of scoring	Quick and easy	Slow, difficult
• Objectivity of scoring	Objective	Subjective
• Scoring criteria	Right/wrong	Rubrics
• Teacher confidence in scoring	High	Possibly low
• Parental acceptance	High	Possibly low
• Comparisons of students	Within and across classes	Usually no norms
• Accommodation of multiple learning styles	No	By design*
• Integration with school grading scheme	High	By design*

*"By design" means that this is true to the extent that it is designed into the assessment.

A2.6,H2

	Multiple Choice	Performance
What is communicated by the assessment		
• Importance of communication as a process skill	If enhanced with open-ended questions	Yes
• Importance of problem-solving skills	May or may not	Yes
• Math and science as a fluid process rather than a static body of facts	No	Yes
Student concerns		
• Good reading skills	Required	Required
• Creativity	Discouraged	Encouraged
• Generation of answers	By teacher or publisher	By student
• Type of questions	Closed	Open ended
• Multiple responses	Discouraged	Encouraged
• Room for explanation	Not allowed	Encouraged
• Choice of answers	Limited; encourages guessing	No limit; guessing more difficult
• One best answer	Implied or required	By design*
• Realistic or relevant tasks	Can be	By design*
• Evidence of subject matter interconnections	Single-subject focus	By design*
• Instructs as well as assesses	No	By design*
• Appearance on paper	Intimidating	White space less intimidating

*"By design" means that this is true to the extent that it is designed into the assessment.

A2.6,H2

Comparison of Multiple Choice and Alternative Assessment #2

***Multiple-choice advantages*:** quick, good for assessing individual skills (e.g., problem solving, critical thinking) in isolation, good for assessing knowledge of facts, can sample a wide range of skills quickly, broad but not deep, cheap

***Multiple-choice disadvantages*:** only one right answer, no opportunity for explanation, hard to measure process skills, doesn't measure ability to use skills to actually solve a real-world problem, assesses skills in isolation, might contribute to the notion that math or science is the memorization of a static body of knowledge, norm-referenced judgments of success

***Performance assessment advantages*:** tasks can be more realistic and holistic, can see process as well as outcome, rubric can promote understanding of the target that transcends this particular problem, the task can be an instructional episode in and of itself, can fit better into a vision of assessment as a tool for learning, can assess skills and abilities hard to assess in multiple-choice format, communicates that problem solving is important, communicates that math is more than the learning of a static body of facts, communicates that math communication is important, criterion-referenced judgments of success

Performance assessment disadvantages: can't sample as well, more labor intensive, takes longer, more expensive, scoring is more judgmental

Summary

	Advantages	Disadvantages
Multiple Choice		
Alternative Assessment		

A2.6,O1

LABORATORY
NETWORK PROGRAM

Activity 2.7

Think of a Time

Purposes:

1. To get participants to think about times in their own lives when assessment was counterproductive or stressful and what made it so; to think about other times when assessment was useful or not stressful and what made it so
2. To change participants' views of what assessment "should" be like
3. To help participants understand why assessment is a difficult topic to talk about and why it is difficult to change

Uses:

1. In Chapter 2 to supplement other "rationale" activities and presentations
2. In Chapter 4 to get participants to think about how the variations in tasks and criteria can affect students
3. In Chapter 5 as part of "consequences"

Rationale:

People have stereotypes in their heads about what assessment "should" be like—filled with anxiety, not fun, hard, high potential for failure, mysterious. This often gets in the way of considering alternatives. This exercise enables participants to explore their preconceptions to determine what they might like to have different.

Materials:

None needed. If you want to summarize the points made, you will need: overhead projector, screen, blank transparencies, transparency pens

Time Required:

30 minutes

Facilitator's Notes:

1. Ask participants to think back to when they were in school, and to recreate the feeling they had the night before a big test. Discuss what it felt like and what made it so.

   Examples of feelings: anxiety, dread, fear, nausea, sleeplessness

   What made it so: being unsure of what would be required, having it be so important, being afraid you would not do well, being afraid to look stupid, having everyone know that you're dumb, facing the consequences with your parents, having to do something that you thought was stupid, the test didn't really measure what you knew.

2. Now ask participants to think back to when they were involved in an assessment that wasn't like this. Discuss what made it so.

   Examples: there was clear and immediate feedback, you knew the criteria for success, the activity was fun and engaging, you could show what you knew in more than one way, you really felt that you had the chance to demonstrate what you knew or could do.

3. If teachers are the participants, ask them to think about the tests they give. What are students' reactions to them? (If desired, summarize points on an overhead.)

Activity 2.8

Sam's Story: Comprehensive Assessment

Purposes:

1. To illustrate the kinds of information that can be obtained from different forms of assessment
2. To demonstrate the importance of using multiple measures to get a more complete picture of a student
3. To demonstrate that each form of assessment has strengths and weaknesses
4. To demonstrate that it is inappropriate to use the results of an assessment intended for one purpose (for example, national comparisons) for another purpose (for example, to make a decision about successful completion of a course).

Rationale:

Participants may have heard rhetoric that suggests that one form of assessment is better than another in all circumstances and for all purposes, and that a single assessment should yield all the information needed. Actually, no single type of assessment can yield all the information needed. Participants will gain perspective on the strengths and weaknesses of various types of assessment by reviewing a student's scores on a norm-referenced test, criterion-referenced test, cognitive abilities test, a psychological review, and a performance task

Materials:

- Overhead projector, screen, blank transparencies, and transparency pens or a flip chart and pen
- Copies of the handouts (**A2.8,H1** to **A2.8,H7**) for each participant, or a comparable set of assessment results for a student at a different grade level (Sam, the subject of this exercise, is a sixth-grader); handouts can also be made into overheads

Time Required:

30 minutes

Facilitator's Notes:

1. The discussion can be done in a small or large group. Hand out copies of each overhead one at a time. Whether in a large group or small groups, have individuals respond to the question at the bottom of the overhead. Repeat for each overhead. Typical questions and comments:

    a. Sam's Scores in Class **(A2.8,H1)** *Are these percents correct? What was covered by each test? Are some scores more important than others?*

    b. Sam's Cognitive Ability Score **(A2.8,H2)** *How much should we take IQ into account when assigning grades? He should be doing good work. What is this test measuring?*

    c. Sam's Norm-Referenced Test (NRT) Scores **(A2.8,H3)** *How does this test relate to what I do in class? Should grades be based on NRT scores? Sam does well with respect to others. Sam needs to be more careful with his computation; his comprehension is good.*

    d. Sam's Criterion-Referenced Test (CRT) Results—Mathematics **(A2.8,H4)** *He needs more work in geometry. His scores are so high here; how does this relate to the lower percentiles on the NRT? This helps planning but not grading.*

    e. An Example of Sam's Problem-Solving Ability **(A2.8,H5)** *Sam can handle multi-step problems, but didn't carry it all the way to the end. Computation was good. One problem is not enough to draw a conclusion. Is this problem the right level of difficulty? How does this compare with the problem-solving score on the CRT? Are any of the tests measuring the same thing?*

2. Wrap up the discussion by asking the group what information each assessment provided and what additional information they needed. Use the final handout to help lead the final discussion **(A2.8,H7)**.

86

SAM'S SCORES IN CLASS

These are Sam's scores in mathematics for the semester. Assign him a grade based on his mastery of mathematics skills, and discuss your instructional plan for him.

100	90	27	68	0
0	89	29	75	89
86	0	92	65	98
100	0	96	72	88

Grade________

Instructional Plan:

What more would you like to know about these scores?

A2.8,H1

SAM'S COGNITIVE ABILITY SCORE

TEST OF COGNITIVE SKILLS

Score = 113

What does this add to your understanding of Sam's performance? How does it support or dispute the grade you assigned him? How would this information cause you to change your grade or your instructional plan?

A2.8,H2

SAM'S NORM-REFERENCED TEST (NRT) SCORES

These are Sam's mathematics scores (percentiles) on the Comprehensive Test of Basic Skills in mathematics.

Total Mathematics	**63**
Math Computation	**53**
Math Concepts and Applications	**71**

What does this add to your understanding of Sam's performance? How does knowing how he performs relative to other sixth graders affect the grade you would assign him? How would knowing about his differential performance in computation and concepts and applications cause you to alter your instructional plan?

A2.8,H3

SAM'S CRITERION-REFERENCED TEST (CRT) RESULTS
MATHEMATICS

State Essential Skills	#Correct	Total Items	Percent Correct	Mastery
1. Problem Solving	11	12	92	+
2. Reasoning	1	1	100	NA*
3. Place Value	6	6	100	+
4. Real Number Sense	4	5	80	+
5. Real Number Computation	18	22	82	+
6. Estimation	9	10	90	+
7. Geometry	2	4	50	-
8. Measurement	10	10	100	+
9. Statistics	9	10	90	+
10. Ratios/Percents Proportions/	7	9	78	+
11. Algebra	4	5	80	+

What does this add to your understanding of Sam's performance? How does knowing how he performs on a number of specific instructional objectives affect the grade you assigned him and the instructional plan you have made for him?

---

*Mastery is determined only if there are at least four items for that skill.

AN EXAMPLE OF SAM'S PROBLEM-SOLVING ABILITY

Michael earns $4.75 per hour and has worked at his job for one month (four weeks at 40 hours per week). He wants to buy a CD player for $125 but needs first to pay $325 for rent, $100 for food, $180 for his car payment, and $125 for his college loan payment. Will he have enough left over to buy his CD? If not, how many more hours does he need to work?

$$\begin{array}{r} 4.75 \\ \times\ 100 \\ \hline 760 \end{array}$$

↖ YES ↘

$$\begin{array}{r} 100 \\ 180 \\ 125 \\ +325 \\ \hline 730 \end{array}$$

How does Sam's performance alter or support your understanding of his mathematical achievement? What additional information does it tell you about his problem-solving strengths and weaknesses that test results did not provide? How might this information help you refine the instructional plan you have created for him?

A2.8,H5

SAM'S PSYCHOLOGICAL ASSESSMENT RESULTS

Sam was assessed by a psychologist, using interview and psychological tests, and the psychologist's findings follow:

> Sam has an above average ability to think analytically.
>
> Sam is a dreamer, has trouble concentrating in class.
>
> Sam is the classic absent-minded professor.
>
> Sam does not like repetitive tasks, nor does he like paying attention to details; it is hard to get him to check his work.
>
> Sam is a "big picture" person.
>
> Sam learns quickly, but when he doesn't understand something, he becomes easily defeated and gives up.
>
> Sam is a born leader in social situations and is well liked.
>
> Sam is kind and compassionate to others, but hard on himself.
>
> Sam is, in general, a well-adjusted boy who doesn't like school, but wants to be successful.

What does this add to your understanding of Sam's performance? What kinds of questions have you been asking yourself about Sam that this information addresses? How would this affect the grade you assigned? How might this information be used to adjust your instructional plan?

A2.8,H6

COMPREHENSIVE ASSESSMENT

To make a decision about Sam's mastery of the mathematics content taught, and to design an instructional plan for him, what additional information do you need to know:

About Sam?

About his math teacher and that teacher's course content, instructional strategies, and grading criteria?

About the tests used to provide these results?

A2.8,H7

Activity 2.9

Going to School

Purposes:

1. To illustrate the importance of clearly defined and well thought-out performance criteria
2. To illustrate different types of performance criteria
3. To discuss the relative advantages and disadvantages of different types of performance criteria

Uses:

1. In Chapter 2 to illustrate the importance of performance criteria
2. In Chapters 4 and 5 to illustrate types of performance criteria and discuss the advantages and disadvantages of each kind

Rationale:

Participants frequently need help in defining "performance criteria," understanding why they are important, seeing the difference between tasks and criteria, understanding the role of performance criteria in instruction and assessment, and understanding the differences between types of criteria and their advantages and disadvantages. This activity is designed as a hands-on way to explore these topics.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Handout A2.9,H1**—*Going to School* (2 pages)
- Two sample student responses to the problem: **Handout A2.9,H2**—*Devon and Mickie* (3 pages)
- Three attached scoring guides: **Handouts A2.9,H3**—*Task Specific Scoring Rubric*, **A2.9,H4**—*Task Specific Points*, and **A2.9,H5**—*Generalized Scoring Rubric—Relative Quality*

- **Overheads A2.9,O1**—*Holistic/Analytical* and **A2.9,O2**—*Generalized/Task Specific*

- Two other scoring guides: holistic descriptive (e.g., **Sample 5.7**—*Mathematics Assessment*) and analytical trait, descriptive (e.g., **Sample 4.2**—*Mathematics Problem Solving*)

- Extra assessment samples if Part C is done: **Sample 4.7**—*Performance Assessment in Mathematics (Alberta)*, **Sample 4.3**—*Discovering the Problem of Solid Waste*, and **Sample 5.1**—*Assessment and Technology Videotape*

Time Required:

30 minutes to 2 hours

Facilitator's Notes:

A. Importance of Performance Criteria (30-45 minutes)

1. Pass out the *Going to School* **(A2.9,H1)** problem and have participants solve it in small groups (10 minutes).

2. Pass out the two student solutions (**A2.9,H2**) and ask participants to "grade them" or "score them" (5 minutes). Participants might ask questions like, "Do you want us to use A-F?" "Do we use a four-point scale?" "Do we take spelling into account?" etc. Just tell them to use their usual procedure. The point of the exercise is to demonstrate that, without definition of how and what to score, it is hard to interpret what the grades or scores are telling us.

3. In a large group, have the various small groups shout out their scores or grades for each student. After they are all recorded, ask the participants to "interpret" them. "What do you know about these students?" (Which they can't do because it is a hodgepodge of numbers and letters based on different scales and criteria.)
   (5-10 minutes)

4. Pass out the criteria packet **(A2.9,H3; A2.9,H4; Sample 4.2; Sample 5.7)**. Ask the participants to score the student response with either the task-specific rubric or the California rubric (or both). (5 minutes for each)

5. Have participants again call out scores and then ask them, "Now what do you know about these students?" (10 minutes) Make sure the following points are made. Rubrics are needed to:

   a. Make it clear to students what it takes to be successful

Activity 2.9

Going to School

Purposes:

1. To illustrate the importance of clearly defined and well thought-out performance criteria
2. To illustrate different types of performance criteria
3. To discuss the relative advantages and disadvantages of different types of performance criteria

Uses:

1. In Chapter 2 to illustrate the importance of performance criteria
2. In Chapters 4 and 5 to illustrate types of performance criteria and discuss the advantages and disadvantages of each kind

Rationale:

Participants frequently need help in defining "performance criteria," understanding why they are important, seeing the difference between tasks and criteria, understanding the role of performance criteria in instruction and assessment, and understanding the differences between types of criteria and their advantages and disadvantages. This activity is designed as a hands-on way to explore these topics.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Handout A2.9,H1**—*Going to School* (2 pages)
- Two sample student responses to the problem: **Handout A2.9,H2**—*Devon and Mickie* (3 pages)
- Three attached scoring guides: **Handouts A2.9,H3**—*Task Specific Scoring Rubric*, **A2.9,H4**—*Task Specific Points*, and **A2.9,H5**—*Generalized Scoring Rubric—Relative Quality*

- **Overheads A2.9,O1**—*Holistic/Analytical* and **A2.9,O2**—*Generalized/Task Specific*
- Two other scoring guides: holistic descriptive (e.g., **Sample 5.7**—*Mathematics Assessment*) and analytical trait, descriptive (e.g., **Sample 4.2**—*Mathematics Problem Solving*)
- Extra assessment samples if Part C is done: **Sample 4.7**—*Performance Assessment in Mathematics (Alberta)*, **Sample 4.3**—*Discovering the Problem of Solid Waste*, and **Sample 5.1**—*Assessment and Technology Videotape*

Time Required:

30 minutes to 2 hours

Facilitator's Notes:

A. Importance of Performance Criteria (30-45 minutes)

1. Pass out the *Going to School* **(A2.9,H1)** problem and have participants solve it in small groups (10 minutes).
2. Pass out the two student solutions (**A2.9,H2**) and ask participants to "grade them" or "score them" (5 minutes). Participants might ask questions like, "Do you want us to use A-F?" "Do we use a four-point scale?" "Do we take spelling into account?" etc. Just tell them to use their usual procedure. The point of the exercise is to demonstrate that, without definition of how and what to score, it is hard to interpret what the grades or scores are telling us.
3. In a large group, have the various small groups shout out their scores or grades for each student. After they are all recorded, ask the participants to "interpret" them. "What do you know about these students?" (Which they can't do because it is a hodgepodge of numbers and letters based on different scales and criteria.)
   (5-10 minutes)
4. Pass out the criteria packet **(A2.9,H3; A2.9,H4; Sample 4.2; Sample 5.7)**. Ask the participants to score the student response with either the task-specific rubric or the California rubric (or both). (5 minutes for each)
5. Have participants again call out scores and then ask them, "Now what do you know about these students?" (10 minutes) Make sure the following points are made. Rubrics are needed to:

   a. Make it clear to students what it takes to be successful

b. Do performance assessment

B. Additional Steps for the Discussion of Relative Advantages and Disadvantages of Different Types of Performance Criteria (40-60 minutes)

Note: Don't do this part of the exercise if you don't feel you can explain these differences.

1. Ask participants to score the student responses with more than one rubric (5 minutes for each; Oregon will take longer).

2. Discuss the differences: holistic vs. analytical trait (**A2.9,O1**), task specific vs. general (**A2.9,O2**), and amount of detail (10 minutes).

3. Discuss the relative advantages and disadvantages of the different types—small group and then large group (20-30 minutes).

TASK-SPECIFIC vs. GENERALIZED criteria:

Advantages of "task specific" and "relative quality": quick, good reliability, fast to learn.

Disadvantages of "task specific" and "relative quality": can't tell what's being assessed by looking at the criteria; an indirect measure of the skill; if the problem is, for example, assessing problem solving, the criteria don't help us to define this concept; doesn't communicate to students what to do differently next time.

Extra disadvantages of "task specific": have to develop a new rubric for each question; some creative student responses that are correct can be scored "wrong" if that option is not listed.

Advantages of detailed, generalized rubrics: increases understanding of what we mean by, for example, good problem solving; if teachers are the raters, their expertise will improve; if students are the scorers, it is an instructional tool; a direct measure of the skill; you can tell what is being assessed by looking at the criteria; different rubrics do not have to be developed for each problem.

Disadvantages of detailed, generalized rubrics: takes longer to learn well.

HOLISTIC vs. ANALYTICAL TRAIT CRITERIA

Advantages of holistic, generalized rubrics (e.g., California): faster to score, easier to learn.

Disadvantages of holistic, generalized rubrics: two students can get the same score for vastly different reasons.

Advantages of analytical trait, generalized rubrics (e.g., Oregon): more diagnostic; more useful for instruction because students can see the relative strengths and weaknesses in their work.

Disadvantages of analytical trait, generalized rubrics: takes longer to learn well.

C. Additional Discussion Topic (10-20 minutes)

How do the various types of performance criteria relate to different visions of the role of assessment in the instructional process? It is important to remember that there is no "right" answer to this question. Different experts have different views on the best way to proceed. Some typical responses:

- Analytical trait, generalized criteria are best for classroom uses, or for large-scale uses where one of the goals is to influence classroom procedures in the maximum way possible.

- Holistic, task-specific, and relative quality criteria might be used to get a quick, overall idea of student performance.

- Generalized criteria are best for the "big" outcomes (problem solving, critical thinking, group collaboration, lifelong learning, etc.) because these occur across tasks and activities and assist students to generalize from one situation to the next. (An excellent example is **Sample 4.7**—*Performance Assessment in Math [Alberta]*.)

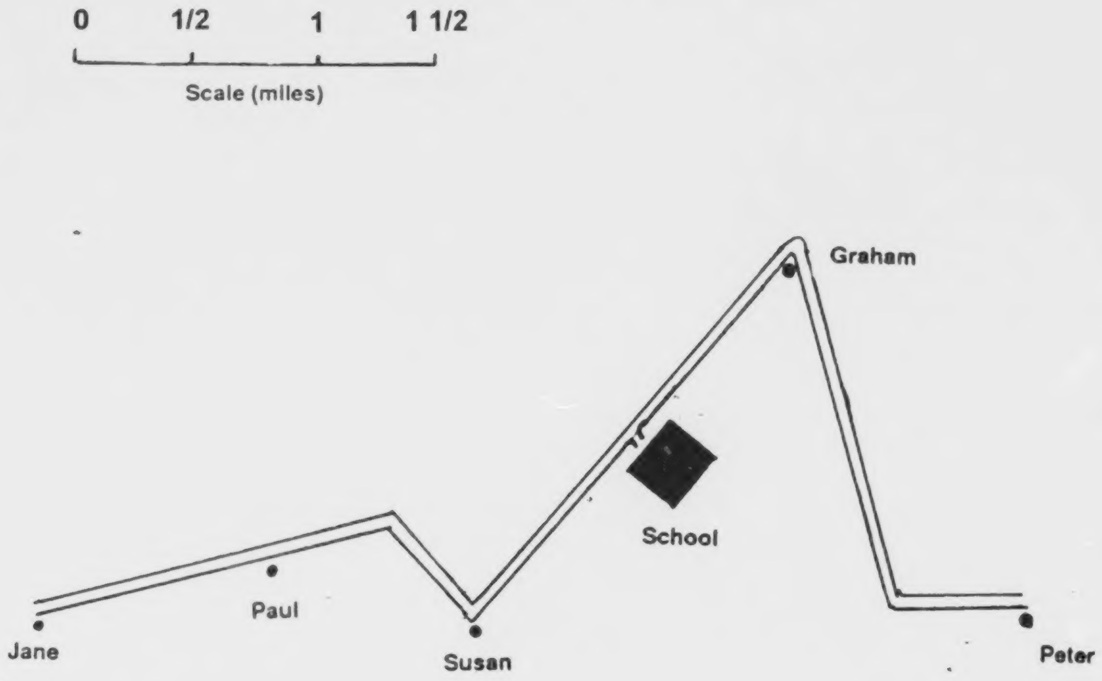
- Task-specific criteria might be acceptable for assessing application of knowledge or skills as in **Sample 4.3**—*Discovering the Problem of Solid Waste*. (However, even here some groups are developing generalized criteria; for example, see **Sample 5.1**—*Assessment and Technology Videotape* for generalized criteria on conceptual understanding.)

- If we see performance assessment as a tool to change interactions between students and teachers in the classroom, then analytical trait, generalized, detailed rubrics are best.

- If we see performance assessment only as a tool to monitor student performance or only as a large-scale accountability tool, then faster procedures might be OK.

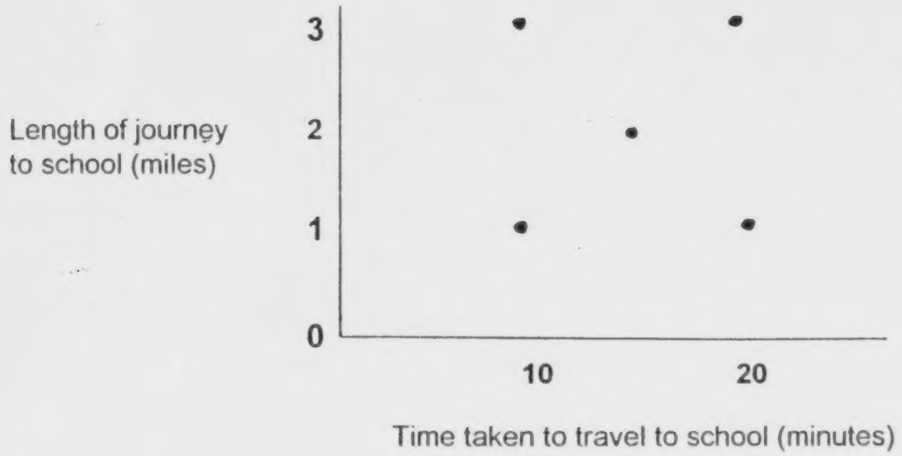
Going to School

Jane, Graham, Susan, Paul and Peter all travel to school along the same country road every morning. Peter goes in his dad's car, Jane rides her bicycle and Susan walks. The other two children vary how they travel from day to day. The map shows where each person lives.


0
1/2
1
1 1/2
Scale (miles)
Graham
School
Paul
Jane
Susan
Peter

Laboratory Network Program—*Alternative Assessment Toolkit*
Activity 2.9—*Going to School*
Author: Judy Arter, NWREL; Glenda Clark and Jackie Palmer, SEDL

This graph describes each pupil's journey to school last Monday.



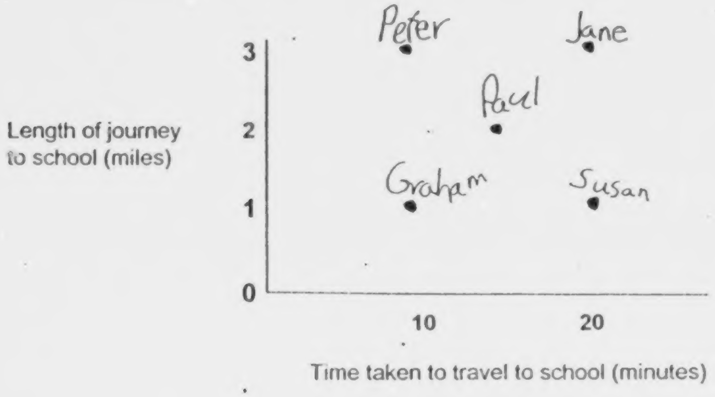
1. Label each point on the graph with the name of the person it represents.
2. How do you think Paul and Graham traveled to school on Monday -- walking, bicycle, or car?

3. Describe how you decided on your answer in question 2.

A2.9,H1

De von

This graph describes each pupil's journey to school last Monday.



1. Label each point on the graph with the name of the person it represents.

2. How do you think Paul and Graham traveled to school on Monday -- walking, bicycle, or car?

I think Graham rode his bike
I think Paul rode his bike

3. Describe how you decided on your answer in question 2.

Paul was 4 miles away from school
and it took him about 30 min. so it
wouldn't be a car because cars are faster over.
and it couldn't be walking be because walking
is to slow.

A2.9,H2

BEST COPY AVAILABLE

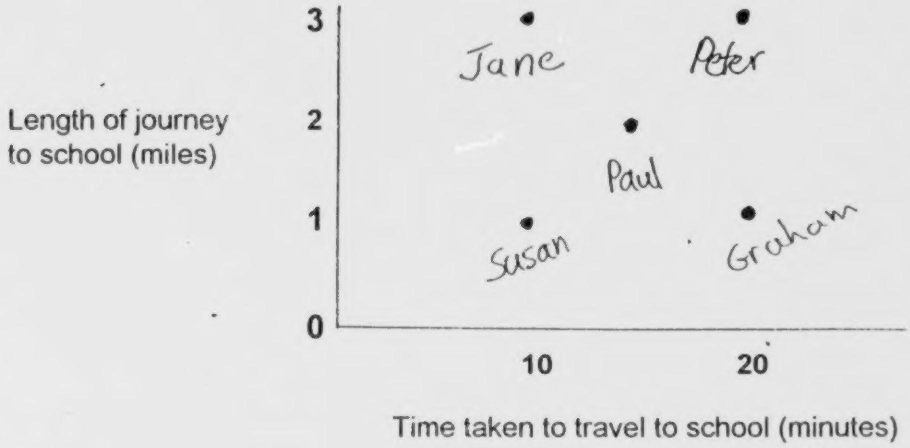
Devon

Graham was 2 miles from school
and it took him 20 min. to get
there. So it couldn't be walking
because it is to school. Driving is
to fast, so it must be bike riding

A2.9,H2

Mickie

This graph describes each pupil's journey to school last Monday.



1. Label each point on the graph with the name of the person it represents.

2. How do you think Paul and Graham traveled to school on Monday -- walking, bicycle, or car?

Graham rode his bike Paul walked

3. Describe how you decided on your answer in question 2.

Graham Lives closer to the schoolthan Peter And Peter rode in his dad's car. He probly rode his bike. Paul is in the midel and he could probly walk.

A2.9,H2

Task-Specific Scoring Rubric "Going to School"

1 Point: Shows no understanding of the requirements of the problem. Is unable to label the graph or give a reasonable answer about how Graham and Paul travel to school.

2 Points: Labels one or two points on the graph correctly. Provides an answer to the type of transportation taken by Graham and/or Paul, although it may be incorrect. Provides some sort of rationale for the answer, although it might include faulty reasoning. For example: "Paul lives between Susan and Jane so he probably rides a bike since Jane does," or "Graham lives the same ways away as Susan so he walks."

3 Points: Has three points on the graph labeled correctly. Has identified "bicycle" as the means of transportation for either Graham or Paul. Provides a reasonable rationale for the choice of answer, although it may not be complete. For example, "Graham is faster than Susan so he takes a bike."

4 Points: Has all or most of the graph labeled correctly (4 out of 5). Has identified "bicycle" as the means of transportation for Graham and Paul. Has an explanation for choice of "bicycle" that shows understanding, but requires some interpretation on the part of the rater. For example, "Paul was in the middle and it took him about 30 minutes. Cars are faster and walking is slower."

5 Points: Has labeled the graph correctly. Has identified "bicycle" as the means of transportation for Graham and Paul and is able to explain how the answer was determined. An example of an excellent explanation is: "Susan lives the same distance away as Graham, and Susan walks. If Graham walked it would probably take about the same time. Since it doesn't take as long he probably rides a bike or in a car. Peter takes a car, but goes a lot further in the same amount of time. Therefore, Graham probably doesn't take a car. So, he takes a bike." The response doesn't necessarily have to have this level of detail to be a "5", but the general line of reasoning needs to be there.

A2.9,H3

Task-Specific Scoring Guide #2
"Going to School"
Points for Various Features

Graph: 1 point for each point labeled correctly (6 points possible)

Graham and Paul: 1 point for correctly identifying that each ride their bike (2 points possible)

Rationale: (4 points possible)

4 points: Excellent rationale—includes accurate comparison to other students

3 points: Good rationale—includes comparison to other students, but some parts might be hard to follow or incorrect

2 points: Fair rationale—Provides a rationale, but includes fairly faulty reasoning

1 point: Poor rationale—shows no understanding of the problem

Generalized Scoring Rubric
Relative Quality—No Explanations

1 Point: Shows no understanding of the requirements of the problem.

2 Points: Shows a little understanding of the requirements of the problem.

3 Points: Shows partial understanding of the requirements of the problem.

4 Points: Shows considerable understanding of the requirements of the problem.

5 Points: Shows complete understanding of the requirements of the problem.

A2.9,H5

HOLISTIC	ANALYTICAL
A single overall score is assigned to a performance task	Several dimensions or traits of a task are scored
Good for large-scale accountability	Assesses relative strengths and weaknesses of students
Quick scoring process	Slower scoring process
Assesses student performance	Diagnostic
Two students can get the same score for different reasons	Assesses changes in student/teacher interactions
Example: California	Example: Oregon

A2.9,O1

LABORATORY NETWORK PROGRAM

Generalized	**Task Specific**
Student ability generalized from performance	Direct measurement of student performance
Statements not specific for a given task	Statements refer to a specific task
Same rubric usable for different tasks	Rubrics must be rewritten for each task
Criteria tell what is being assessed	Criteria do not tell what is being assessed
Can assess "big" outcomes, e.g., problem solving or critical thinking	Can assess application of knowledge and skills
Good student instructional tool	Quick to administer and score
	Reliable

A2.9,O2

LABORATORY
NETWORK PROGRAM

Activity 2.10

Tasks vs. Performance Criteria

Purpose:

To illustrate the difference between tasks and performance criteria

Uses:

1. In Chapter 2 as an activity that introduces participants to alternative assessment

2. In Chapters 4 and 5—when showing samples it helps people understand the variations you are about to show; for example, this could be a prelude to **Activity 2.9**—*Going to School*, **Activity 4.1**—*Performance Tasks—Keys to Success*, or **Activity 4.3**—*Performance Criteria—Keys to Success*

3. In Chapters 4 and 5 as a prelude to the development or selection of an alternative assessment so that participants know they have to attend to both tasks and criteria; for example, this could be a prelude to **Activity 5.1**—*How to Be a Good Consumer*

Rationale:

Participants are often confused about what alternative assessments are. This activity not only shows them examples, it also helps them to understand that there are two major parts of an alternative assessment—the task and the performance criteria. Many people, when they think of assessment, only think of the task. But, without criteria, there is no assessment.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens

- Three examples of tasks and related performance criteria; a good variety is: **Sample 4.1**—*How Many Buttons?* (*task*=grade four, math project, with equipment and group work; *criteria*=task specific), **Sample 4.9**—*Mapping the Blue Part* (*task*=grade eight, science on-demand performance assessment, paper and pencil, individual work; *criteria*=generalized criteria tailored to specific tasks), and **Sample 5.3**—*Science Portfolio (GSE)* (*task*=grade 12, assemble a science portfolio and write reflections; *criteria*=holistic, one for each project, generalized)

Time Required:

20 minutes

Facilitator's Notes:

1. This can be a demonstration with all examples on the overhead, or participants can have the examples as handouts. I have also used abridged versions on the overhead and expanded versions in the handouts (and vice versa).

2. The presenter can tell the participants everything, or the participants can look through the examples in small groups and then have a large-group discussion. The points to make are:

   a. If this is a kick-off activity, you will need to define "criteria" and "task." Ask:

   What do you think of when you hear the word "criteria"? (Typical responses: standards, what I value, keys to success) *What are our criteria for a restaurant?* (Typical responses: cleanliness, price, quality of the food, service, convenience, nonsmoking section) *What are our criteria for math problem solving?* (Participants usually have more difficulty with academic criteria. Use this as a segue into the notion that we need criteria in order to judge the performances in an alternative assessment.)

   Tasks are the activities that we give students to do (e.g., solve a problem, do a dive, run a restaurant); criteria are the means by which we judge the quality of the response or performance. Ask the participants to generate examples of tasks.

   b. An alternative assessment is more than just a task. One has to plan both the task and the scoring mechanism.

   c. Tasks and criteria vary. Some of the differences are illustrated in the samples. For example, tasks can differ by time, complexity (constructed response, on-demand, project, or portfolio), written or manipulative, group or individual. Criteria can vary by task specificity, amount of detail, one score or more than one score. (Other exercises will encourage participants to discuss the advantages and disadvantages of these differences.)

Chapter 3

Integrating Assessment with Curriculum and Instruction

What's in this Chapter?

This chapter considers the ways assessment influences teachers, instruction, and students. Such consideration is essential if we want to build an assessment system based on a clear vision of what we want assessment to accomplish. The technical term related to the need to look at the consequences of our assessment activities is "consequential validity." One of the lessons learned from past experiences with assessment is that looking at traditional notions of validity (i.e., Does the test measure what is intended?) is not enough. An additional validity consideration is the extent to which an assessment has positive consequences for the educational enterprise. One of the criticisms of past assessment activities, for example, is that multiple-choice tests narrow instruction and reinforce the notion that everything has a "right" answer and there is only one right answer. Nothing about alternative assessment automatically ensures that such negative unintended side effects won't occur. Therefore, it behooves us to consider these issues, as much as possible, in advance.

Chapter Goals

1. Present different conceptions of what it means to integrate assessment and instruction
2. Discuss various ways that assessment can influence teachers and students
3. Through these discussions assist the reader to build a vision of what he or she would like assessment to accomplish
4. Discuss the assessment design implications of various visions

Chapter Content

Readings

Views of Integrating Assessment with Curriculum and Instruction .... Page 4

The first section discusses various ways that alternative assessment can influence curriculum, instruction, teachers, and students. This section will assist the reader with developing a vision of what he or she would like assessment to accomplish.

Continuous Monitoring .... Page 9

The second section expands on one of the views in the first section—the idea of assessment being a tool for continuously monitoring student achievement and progress so that instruction can be rationally planned.

Using Assessment as a Tool for Learning .... Page 21

The third section expands on another of the views in the first section—the idea that alternative assessments can be more than a way for tracking student progress. If designed properly, they can also be tools for learning. That is, student achievement is enhanced in the process of assessing and being assessed.

References .... Page 29

Activities in this Chapter

Activity 3.1 *Camping Trip: A Mathematics Problem-Solving Task*

This activity involves sorting samples of grade five student work into three stacks representing "strong," "medium," and "weak" responses to a problem-solving activity. Participants have the opportunity to describe and discuss the characteristics of work that differentiate these stacks. This is an exercise in developing performance criteria, but is also an exercise that demonstrates how developing performance criteria can help sharpen the goals we have for students, increase teacher expertise, and assist students with understanding what to do. Time: 60-90 minutes

Activity 3.2 *Sorting Fish: A Science Task*

This is similar to **Activity 3.1**, but uses a science task instead of a mathematics task. Time: 60 minutes

Activity 3.3 *Miss Tolivar's Mathematics Class*

This activity uses a video that shows a teacher implementing continuous monitoring of student achievement. Time: 60 minutes

Related Activities in Other Chapters

Activity 6.2 *How Can We Know They're Learning?*

A version of **Activities 3.1** and **3.2** designed for parents. Time: 2 hours

Sample Assessments Included in this Chapter

Figures 3.3-3.6 *Fox Rot*

This is a computer-based assessment making use of interactive multimedia. It illustrates the points made in the section on "Continuous Monitoring."

Sample Assessments Referred to in this Chapter

One additional sample assessment is used to illustrate some of the points made in this chapter. The sample itself is found in **Appendix A**.

Sample 3.1 *Acid-Bases Assessment Task—Cleaning Up a Toxic Spill*

Other Resources

Build Your Own Performance Criteria

Outlines the steps in developing generalized performance criteria using a running example of student self-reflection. This outline is the last three pages of **Activity 3.1.**

Integrating Assessment with Curriculum and Instruction —Readings

Views of Integrating Assessment with Curriculum and Instruction

This section discusses various ways that alternative assessment can influence curriculum, instruction, teachers, and students. Such influences can also be thought of as different approaches to integrating assessment and instruction. This section will assist the reader in developing a vision of what he or she would like assessment to accomplish by looking at four approaches to integrating assessment and instruction—continuous monitoring, pushing instruction, assessment development, and tools for learning.

Assessment as Continuous Monitoring

Assessment can be part of an instructional feedback loop that helps us monitor student progress toward those goals we have for them. The monitoring notion applies equally to large-scale and classroom assessment. Monitoring is used at the large-scale level for making decisions about the quality of education students are receiving. At the classroom level, teachers need information about the status of student progress toward goals in order to effectively plan what to do next. The teacher who continuously monitors can be contrasted with the teacher who teaches an entire unit and then assesses, teaches a unit and then assesses, etc.

This is not a new concept. It is a very important purpose of assessment that we have long used. What might be new are: (1) the types of student outcomes on which we are monitoring progress (more emphasis on process skills and the "big" outcomes of collaboration, critical thinking, taking control of learning, etc.); (2) the type of assessments we need to use to monitor progress on these outcomes (e.g., more performance-based); (3) more dependence on teachers to collect the needed information for large-scale assessment; and (4) the need, at the classroom level, to collect more systematic information (intuition no longer is enough).

The position that assessment provides the information that powers instructional decisionmaking is stated nicely in the second section of this chapter. This section describes new thinking about science and mathematics instruction and provides examples of how alternative assessment information fits into the instructional cycle at the classroom level. The theme of continuous monitoring is illustrated in **Activity 3.3**—*Miss Tolivar's Mathematics Class.*

The next three views of "integrating assessment and instruction" involve a more proactive role for assessment.

Assessment as a Tool to Push Instruction in a Desired Direction

This view of integrating assessment and instruction is based on the position that large-scale assessment communicates to teachers what and how to teach. Therefore, if we design our large-scale assessments correctly, we will drive instruction in the desired direction. Gong, Venezky and Mioduser (1992) describe a study done in science classrooms in which performance assessment tasks were explicitly designed to model current theories of good instruction—long-term projects, clear outcomes, emphasis on the process of science, student collaboration, etc. The authors argued that this type of modeling is necessary because many teachers don't know how to teach science in order to attain the types of goals currently being proposed for students. They found an impact on teachers in terms of their (1) perception of science, (2) knowledge of science, (3) awareness of students, (4) teaching styles and methods, and (5) professional interaction.

Shavelson and Baxter (1992) provide another good statement of this view. They describe how current knowledge about how students learn and the changing nature of our goals for students require different instructional techniques, classroom configuration, and assessment. Since teachers "teach to the test," if we design certain features into our assessments, then teachers will *have to* alter their instruction if only to ensure that their students do well on the test. Shavelson and Baxter specifically discuss current knowledge about how students learn, the changing nature of our goals for students in science and mathematics, and, as a result, features that need to be designed into assessments.

Please note that this view emphasizes the impact of assessment *tasks* on instruction. Teachers model subsequent instructional activities on the activities/tasks in the assessment. For example, once teachers see how to incorporate collaboration into an activity, they apply it to later activities. This supplements other views in this section in which it is the *criteria* for judging performance that greatly impacts instruction.

Assessment *Development* as a Strategy to Change Instruction

This view is similar to the previous one, except that it is the *development* of the assessment (in addition to the use of the assessment) that changes instructional practice. When writing about their experiences in trying to develop meaningful alternative assessments, author after author expresses the opinion that the process of developing the assessment is at least as useful as having the assessment in place at the end. Statements such as the following are typical:

> *The desired outcome is fully realized only when the participants themselves work through the issues critical to achieving it. The processes of exploration and search are important components in changing educational practice; it is not enough to pass on the wisdom of others' experience.* (Murphy & Smith, 1990)
>
> *The dialogue that is happening among teachers is one of the most exciting things happening from this. Portfolios may actually bring us together more than ever before so we begin to define our criteria much more precisely.* (Sugarman, 1989)
>
> *The benefits of portfolios lie as much in the discussions they generate among teachers—and between teachers and students—as in the wealth of information they provide.* (Murphy and Smith, 1990)
>
> *To our surprise, the [model student papers we chose to illustrate points on the scoring guide] were just one benefit of this entire process. More important, we felt, were the discussion and negotiation that occurred that afternoon. In addition to learning about student writing, we valued the opportunity to talk to one another about the work we do. That sense of being a community of learners continued for the entire year.* (Evans, 1993)

In order to develop a meaningful alternative assessment, one has to have expertise in the content area to be assessed, in how learning occurs, and in assessment. The person who develops the assessment does the learning. Why not make these developers teachers so that they can take this increase in expertise back to the classroom? Costa and Kallick (in press) are very clear on this point. If we want meaningful change in assessment, we need to re-enfranchise teachers and students by giving them the power to assess progress. To accomplish this, we need to support schools in becoming learning organizations and put teachers back in control of developing assessments.

Please note that in this view, it is the development of the performance *criteria* rather than the performance *tasks* that is frequently cited as having this positive impact on teaching expertise.

Assessment as a Tool for Learning

A fourth view of integrating assessment and instruction is based on the view that assessment activities *themselves* (both large-scale and classroom) influence students—for better or worse. In this view, assessment is not a neutral episode of collecting information (as in view #1), nor is it merely a way of influencing future instruction by teachers (as in views #2 and #3), but it is, itself, a way of influencing students. If designed properly, students can be influenced in positive ways. For example, students could actually learn something from doing the assessment—the assessment activity itself *is* an instructional episode. In other words, assessment activities not only monitor student progress toward goals, but they can actually *help us accomplish* the goals.

One view of how this influence works is nicely presented by Stayter and Johnston (1990). They detail the many ways that students get messages from our assessments—what to study, who has control of judgment, student role in the classroom, etc. They also provide specific examples of how to design assessments to send productive messages. Although their examples come from the area of literacy assessment, their comments apply to other areas as well. **Activity 5.2**—*Dr. Know-It-All* provides a light-hearted way for participants to explore some of the messages our assessments send.

Another slant on the influence of assessment on students is articulated by the third section in this chapter, which uses examples in mathematics to show how clearly articulated performance criteria can promote student learning. This section asserts that assessments are learning episodes to the extent that they help teachers and students to clarify the expected outcome of instruction and promote student self-evaluation and reflection. In this view, alternative assessments influence students through use of the *performance criteria*. This power is illustrated in **Activity 3.1**—*Camping Trip,* **Activity 3.2**—*Sorting Fish,* and **Activity 6.2**—*How Can We Know They're Learning?*

A third way that assessments can directly influence students is illustrated by Gong, et al. (1992) who describe how the *tasks* used in their assessment were instructional—students learned from the extended projects that were the basis for the assessment. **Sample 3.1**—*Acid-Bases Assessment Task* provides an example of an assessment task that is also instructional.

In summary, alternative assessments can have an impact on students through the messages they send, through appropriate use of performance criteria, and from completing the assessment task itself.

Implications of Assessment Design and Development

Although it is interesting and informative to consider the various views of "alignment" and "integrating assessment with instruction," it is also imperative to do so. One's philosophy and unstated assumptions about "integration" will affect how alternative assessments are designed and developed. For example, if *you* would like *your* assessment to push instruction in a new direction, then not only must you enlist teachers in the development, but you must have a clear idea of the specific instructional changes that are needed so that they can be designed into the assessment.

Likewise, if *you* want to be able to use assessment as a tool for learning, you must design certain features into the assessment. For example, in order to increase the instructional usefulness of alternative assessments, performance criteria should be designed to be generalizable across tasks—they can be used to judge, for example, many student labs and not just the current lab. If we design our scoring procedures to be task specific, students will not be as likely to be able to generalize them to the next task. If our goal is to create students who "can win again," what good does it do to create mechanisms for them to see how to "win" only on the current task?

Take some time to think about *your* philosophy with respect to integrating assessment and instruction to determine implications for assessment design. Chapters 4 and 5 will also assist in this process.

Continuous Monitoring

Why Integrate Assessment and Instruction?

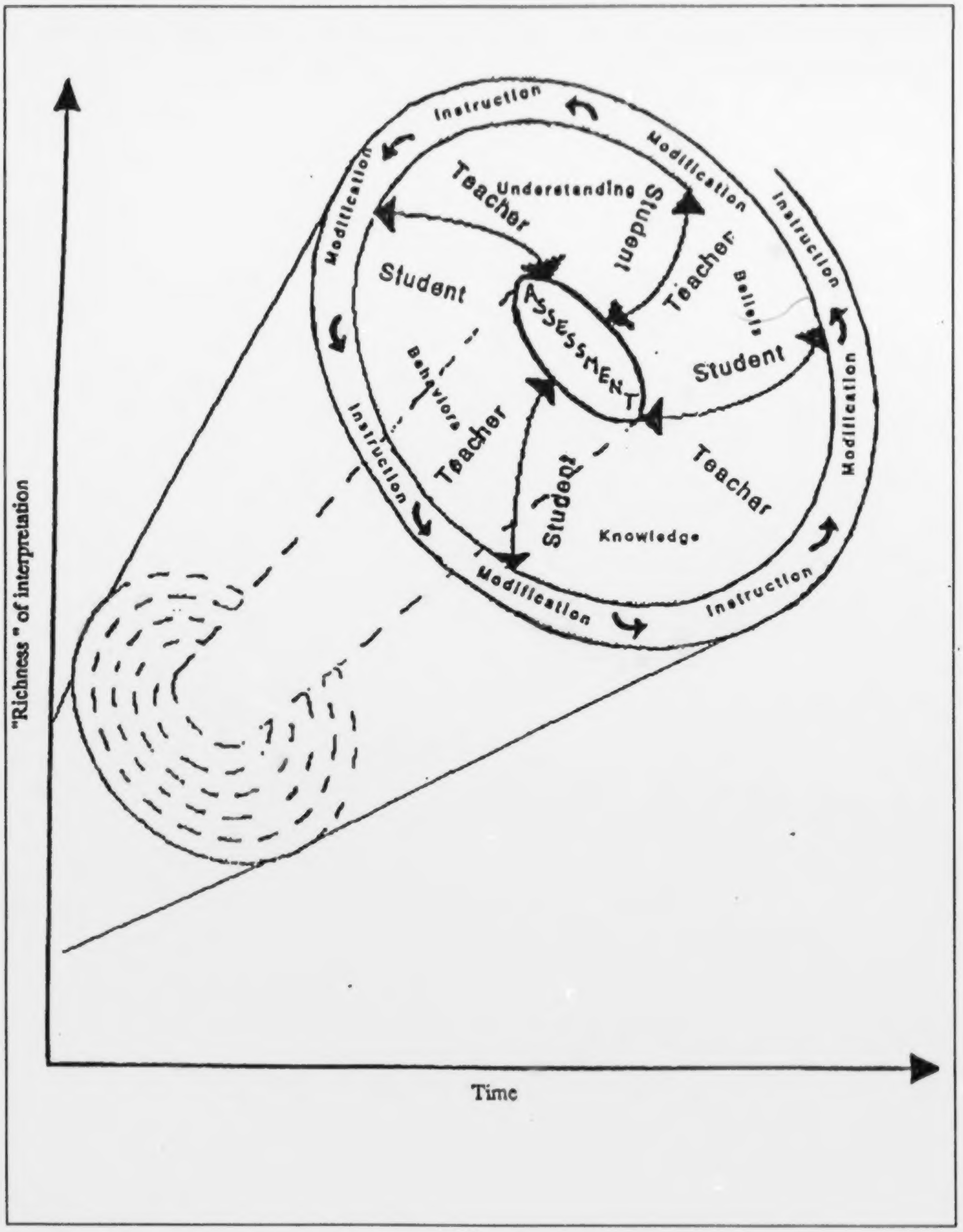
> *From today's cognitive perspective, meaningful learning is reflective, constructive, and self-regulated.... People do not merely record factual information but create their own unique understandings of the world—their own knowledge structures. To know something is not just to passively receive information, but to interpret it and incorporate it into one's prior knowledge. In addition, we now recognize the importance of knowing not just how to perform, but also when to perform and how to adapt that performance to new situations. The presence or absence of discrete bits of information, which is typically the focus of many traditional multiple-choice tests, is not of primary importance in the assessment of meaningful learning. Instead, we care more about how and whether students organize, structure, and use that information in context to solve problems.* (Herman, Aschbacher, and Winters, 1992, pp. 14-15)

Based on the cognitive psychology of how learning takes place in individuals, it is imperative for the teacher to constantly monitor the child in order to meet the needs of that child. Spector (1993) describes the new instructional paradigm as one that is generative (guided by student questioning), inclusive (integrating multiple cultural, racial, and gender values) and constructivist (building on student's biological and experiential prior knowledge). It mirrors good science, establishes links, and supports active, experiential learning. It fosters collaboration and takes into account individual learning styles and stages of cognitive development.

Similarly, the role of assessment in the new paradigm must also shift. Rather than rewarding only correct responses, assessment must inquire about the reasons for obtaining incorrect responses (which may not, on reflection, be incorrect at all, if viewed from the student's perspective). The role shift emphasizes different kinds of assessment for different purposes, and multiple forms of assessment for determining student learning. Assessment is geared to improving instruction, to stimulating inquiry and personal growth in students, and to fostering cooperative learning. In short, it must assess what society values, and in the educational arena, that is student learning.

The ideal approach for constant monitoring of student learning is the integration of assessment and instruction. In its ultimate form, this integration is so complete that the lines between assessment and instruction fade. The student is completely unaware of being assessed, of instruction being modified on the spot, and of further cycles of assessment, instruction, modification of instruction, re-assessment, ad infinitum. In the new educational paradigm, then, the integration of assessment and instruction can be viewed as a spiral. Figure 3.1, based on a constructivist model of learning, depicts this spiral approach.*

Figure 3.1
A Slice of Time in a Constructivist Continuum of Curriculum Development



Instruction
Modification
Instruction
Modification
Instruction
Modification
Instruction
Modification
Teacher
Understanding
Student
Student
Teacher
Beliefs
Assessment
Student
Behaviors
Teacher
Student
Teacher
Knowledge
"Richness" of interpretation
Time

---

* Thanks to Dr. Tom Peters who suggested the three-dimensional approach.

In this model, assessment is an ongoing activity in the instruction cycle (process) designed to create an optimal learning situation for students. It results in continual evaluation and adjustment by the teacher. For example, when students are asked to explain how they arrived at their answer to a problem calling for a comparison of the area of a field to its perimeter, the teacher is able to assess student understanding of the concepts of area and perimeter, ratio, computation, and non-mathematical skills such as reading and comprehension. If a student is not successful in solving the problem because of misconceptions about the meaning of ratio, the teacher can provide experiences that enable the student to master that concept. The teacher might also determine through this informal assessment that students have already mastered a concept that was planned for instruction in the future, such as symmetry, and eliminate that instruction.

As this cycle of assessment/instruction is conducted by the teacher, adjustments are made that reflect information gained, learning is reassessed, and further adjustments are made as required by new assessment information. In the role of teacher-practitioner, the teacher becomes more knowledgeable about pedagogy and classroom application, which results in growth of the individual teacher to deliver appropriate instruction to all children. The assessment becomes, therefore, not an end in itself, but a tool for discerning the need for change.

Effective instruction *must* be integrated with assessment. To fail to do so means to continue to lose from the educational pipeline those students who have traditionally not performed well in the old educational paradigm, e.g., those "at risk." *Not* integrating assessment and instruction substantially hinders progress towards achieving the third National Educational Goal put forth by President Bush and the National Governors Association: "...every school in America will ensure that *all* students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy" (italics added) (*America 2000*, 1991).

What Is Integration of Assessment and Instruction When Viewed as Continuous Monitoring?

In general, newer science and mathematics curriculum materials advocate assessments that:

- Are embedded within instructional materials
- Use a variety of methods to assess the student's progress
- Emphasize teacher observation and teacher judgment
- Provide methods for getting at the reasons behind children's answers (Perrone, 1991, p. 18)

Having established the efficacy of integrating assessment and instruction, there is still much confusion about what integrated assessment and instruction looks like. *There are basically two key characteristics of integrated assessment and instruction: first, integrated assessment must be a process of inquiry by the teacher, and second, the assessment must impact the instructional process and thus contribute to learning.*

Continuous Assessment of Student Learning/Developmental Growth

To get a more concrete view of successfully integrated assessment and instruction, it is helpful to review the following list of assessment methods modified from Perrone (1991):

- Observation
- Verbal response (interviews, oral examinations, informal oral questioning)
- Written records, narrative form (essays, stories, descriptions of processes)
- Written records, drawings/illustrations (diagrams, conceptual maps and webs, graphs, sketches, detailed drawings)
- Demonstration of skills (dramatic responses, successful experiments, use of manipulatives)
- Products (models, displays, projects, dramatic presentations)

Note that many of these methods can be used in tandem. For instance, the **student-as-teacher** approach to integrating assessment and instruction may utilize all of the above approaches. In one integrated assessment/instruction task developed by a high school teacher, for example, senior physics students use an inflatable portable planetarium to learn and teach basic astronomy facts. In pairs, they research and prepare an astronomy teaching presentation for one or more science classes in the school. A teacher-developed test is used to determine astronomy content knowledge in order to re-teach concepts so that students feel comfortable teaching them. Using various teacher-developed rubrics, students are continually assessed on their research and collaborative skills. During planetarium productions, they are further assessed on their responses to student questions and adaptations of their presentation to the audience—skills particularly enhanced by content knowledge and understanding.

In another student-as-teacher approach that also combines **self- and peer assessment**, a team of high school science teachers developed an integrated assessment and instruction activity in which pairs of students prepare chemistry demonstrations focusing on selected biological and chemical processes. They then perform the demonstrations for younger students in other science classes, responding to their questions, while class members videotape the productions. Class members subsequently critique their own and each other's performances using the videotapes and criteria developed jointly by the students and teacher.

Figure 3.2, an excerpt from a paper on **cognitively guided instruction** (Fennema, Carpenter, and Peterson, 1989), further demonstrates how assessment and instruction can be effectively integrated in mathematics.

Figure 3.2
Excerpt from "Learning Mathematics with Understanding"

A major problem in the sequence of development of mature problem solutions is the use of symbols. Consider how Ms. Miller helped her children to develop their own understanding of the complex process of writing number sentences.

> *The children in Ms. Miller's class could solve "join result-unknown" problems and "separate result-unknown" problems quite well. Some children used direct modeling, some used counting on, and a few, derived facts. Some children recalled facts with small numbers. No one used written symbols. Ms. Miller asked a small group to join her near the blackboard and asked them to solve a "join result-unknown." Ms. Miller then showed the children how that problem looked when represented by symbols and asked the children to explain the symbols. One of the children gave the following explanation for transferring what she did with her counters to the number sentence. "Numbers show the same things the counters show." Ms. Miller continued to have the children solve problems and to write the number sentences on the board. She always asked the children to relate how they had solved the problem to the number sentences.*

Ms. Miller had observed her children and decided that this group was ready to move on to using symbols in conjunction with their more primitive problem-solving procedures. For the next several months, she carefully monitored their understanding of the relationships between strategies such as direct modeling and the writing of symbols. She never expected children to perform operations with symbols that they could not perform readily by direct modeling or counting strategies. However, she recognized that the most mature strategy involved symbols, so she continually pointed out the relationship of symbols to their less mature strategies. She believed that the children would eventually construct the relationship between symbols and other strategies and thus come to use the symbols with understanding.

--Fennema, Carpenter, and Peterson, 1989
(permission to use granted by Fennema)

Note in this example that the assessment is a continual process aimed at identifying the students' cognitive level so that instruction can be modified accordingly. Without the continuous assessment, the instructor would be ignorant of whether or not students were cognitively ready to move on to the next stage of understanding. The **cognitively guided instruction model** thus has both of the key characteristics of integrated assessment and instruction—it is a circular process of inquiry used to modify instruction.

The Mathematical Sciences Education Board, in the text *Measuring Up: Prototypes for Mathematics Assessment* (1993), provides prototypes of integrated instructional/assessment activities that support the NCTM Standards. In the prototype "Bridges," elementary students are provided with **manipulatives** and student worksheets with the following instructions:

> *You are an engineer who wants to build different kinds of bridges. The bridges will be made of colored rods. The first bridge you are to build is a 1-span bridge made with one yellow rod and two red rods... The yellow rod is called a span and the red rods are called supports. Since the yellow rod is 5 cm long, the length of the bridge is 5 cm... As you build bridges in the following activities, think of a way to keep track of the number of rods of different colors you use. Your goal is to find out how many rods of each color you would need to build a bridge of any size.* "Measuring Up," 1993, p. 44

Students then proceed to work through the activity. They are encouraged to monitor their own understanding and to "decide for themselves what they need to do to understand the situation" (p. 49). The task, while focusing on "the generation and identification of patterns in a numerical setting...is designed so that a variety of approaches to the computations are feasible.... More important than the computation itself, however, are the students' decisions about what computations are appropriate, and the connection between the computation and the dimensions of the bridge" (p. 49). It is these ongoing decisions that must be monitored with ongoing assessment.

The advent of new technologies has the potential to further expand our ability to integrate assessment and instruction. An example of a prototype instrument currently under development is provided in the LNP Alternative Assessment database.* The *Multimedia-Based Assessment Sampler* uses a computer linked to a video disc player to simultaneously instruct and assess student process skills.** Eleven performance tasks are being piloted at both the elementary and middle school levels. In the task "The Fox Rot" (see Figures 3.3, 3.4, 3.5, and 3.6), students observe photos of a fox decomposing, place them in chronological order, and explain reasons for their ordering. Students may then view a short movie sequence as a form of self-evaluation, after which they can modify their responses. Responses may either be elicited verbally by the teacher, or they can be entered into a CLARIS Works program for printing or electronic network access by the teacher. Thus, assessment is by the student and/or teacher, with immediate feedback and modification of instruction readily administered.

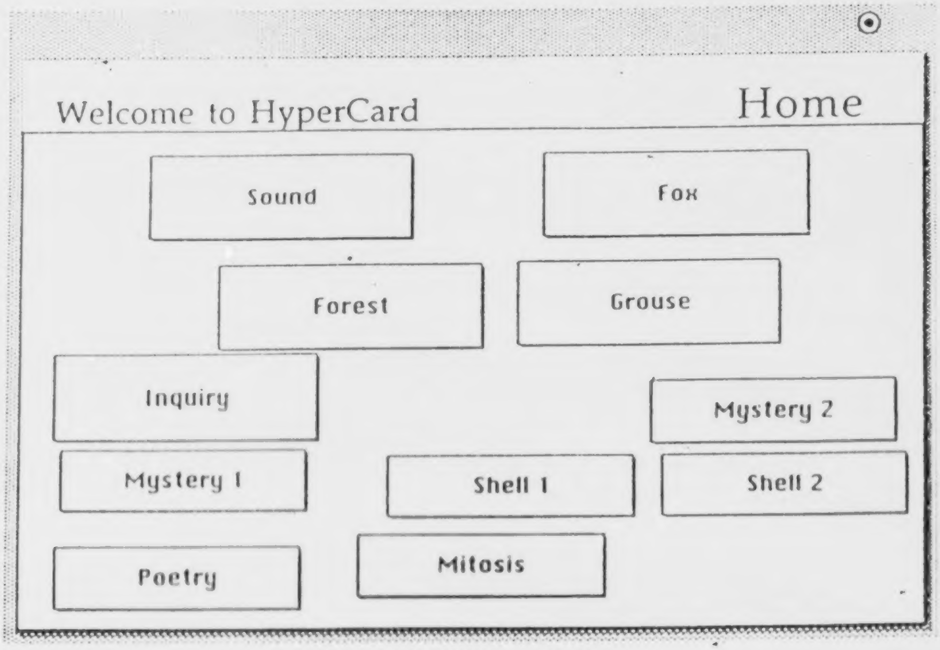
---

* The LNP-AA database is a compendium of descriptive information on 199 alternative assessments in science and mathematics. It is available from Document Reproduction Service, Northwest Regional Educational Laboratory, 101 SW Main St., Suite 500, Portland, OR 97204, (503) 275-9519. (NOTE: This database is available for both IBM and Macintosh computers and uses FileMaker Pro software.)

From each of the above examples, key words have been highlighted to indicate approaches that have been successfully employed to integrate assessment and instruction as a monitoring tool: student-as-teacher, self-assessment, peer assessment, **cognitively guided instruction**, manipulation of objects to assist in making mental connections, and application of new technologies for integrating assessment and instruction. This is not, by any means, an exclusive list of approaches. Any of the tasks listed at the beginning of this section can serve, either singly or in combination, as vehicles for integrating assessment and instruction. They need only be applied with the purpose of inquiring about the status and/or process of student learning, with the intention of modifying instruction based on that assessment, in a continuous fashion.

** A *Multimedia-Based Assessment Sampler* is available from the AOP Hub/Midwestern State University Assessment Consortium. The *Sampler* includes portions of the most recent field test version of "The Fox Rot" along with other tasks under development. The *Sampler* is formatted for Macintosh computers only. (4MB RAM, a hard drive, and a color monitor are recommended. Quick Time and Hyper Card 2.1 software are required.) To receive the *Sampler*, send a 3.5" high-density (1MB or greater) disk and $2.00 for postage and handling to the author (see reference list at end of this chapter).

Figure 3.3
Home Card and Assessment Task Information



"Non Enriched" Performance Task: an activity designed to assess a narrow range of student achievement which can be completed in a short time frame (one class period or less).

Key Features:

prompt: a charge to student to undertake a task which leads to a performance or product.

rubric: a scoring system enabling the assessor to place value on important aspects of the performance or product. (narrows the range of achievements measured)

Purpose:

formative: setting value on student achievement during an instructional episode. The primary purpose is to direct future instruction.

DOMAIN II - Exploring & Discovering. Processes of Science.
DOMAIN III - Imagining & Creating. Higher Order Thinking/Problem Solving

–Courtesy of AOP Hub/Midwestern State University Assessment Consortium, "Multimedia-Based Assessment Sampler," (Work in progress. Video footage from Optical Data's "Windows on Science," Texas Edition, 1990).

Figure 3.4
Initial Prompt and Rubric - The Fox Rot

Performance Task—The Fox Rot

Part 1: Field Research

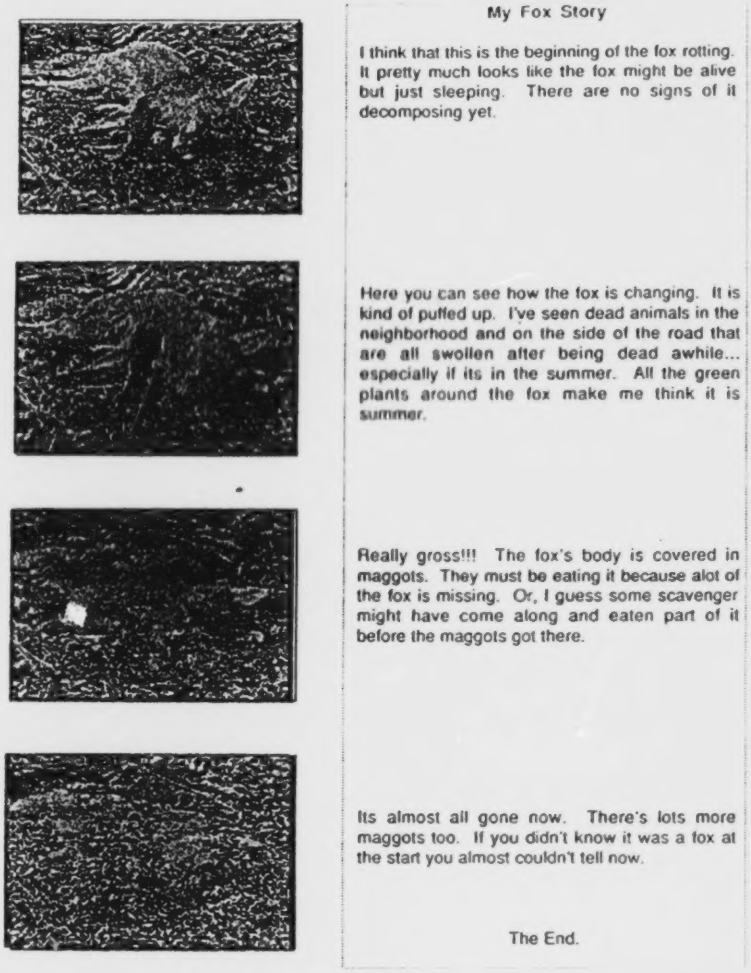
PROMPT: Observe the 4 photos of a fox decomposing. Place the photos in order from start of decomposition (top) to end of decomposition (bottom). List the observations that you make to support the order in which you place each photo. Be sure to compare data from each photo.

RUBRIC - Observing/Inferring:
TX Essential Element 2 (B, C, D) – Acquire data through senses.
TX Essential Element 4 (B, D) – Communicate data and information.
TX Essential Element 6 (B) – Draw logical inferences.

POINTS	CHARACTERISTICS
4	Observes/infers properties of the fox (which may include its surrounding environment) for all four photos.
3	Observes/infers properties of the fox (which may include its surrounding environment) for three photos.
2	Observes/infers properties of the fox (which may include its surrounding environment) for two photos.
1	Observes/infers properties of the fox (which may include its surrounding environment) for one photo.
0	Makes no observations/inferences.

–Courtesy of AOP Hub/Midwestern State University Assessment Consortium, "Multimedia-Based Assessment Sampler," (Work in progress. Video footage from Optical Data's "Windows on Science," Texas Edition, 1990).

Figure 3.5
Sample Student Response



My Fox Story

I think that this is the beginning of the fox rotting. It pretty much looks like the fox might be alive but just sleeping. There are no signs of it decomposing yet.

Here you can see how the fox is changing. It is kind of puffed up. I've seen dead animals in the neighborhood and on the side of the road that are all swollen after being dead awhile... especially if its in the summer. All the green plants around the fox make me think it is summer.

Really gross!!! The fox's body is covered in maggots. They must be eating it because alot of the fox is missing. Or, I guess some scavenger might have come along and eaten part of it before the maggots got there.

Its almost all gone now. There's lots more maggots too. If you didn't know it was a fox at the start you almost couldn't tell now.

The End.

–Courtesy of AOP Hub/Midwestern State University Assessment Consortium, "Multimedia-Based Assessment Sampler," (Work in progress. Video footage from Optical Data's "Windows on Science," Texas Edition, 1990).

Figure 3.6
Verification Prompt and Rubrics

Part 2: Verification

PROMPT: View the time lapse sequence of a fox decomposing. You may view it as often as you like. Compare the sequence of photos you made in Part 1 with the time lapse sequence. Does your photo order match the time lapse sequence? Explain your answer using observations to support you conclusion.

RUBRIC - Conclusions:
TX Essential Element 4 (F) – interpret the arrangement of data on...visuals
TX Essential Element 6 (D) – draw conclusions from observed data.

POINTS	CHARACTERISTICS
4	Draws conclusion. Bases conclusion on observations comparing all still photos of the fox with the time lapse segment.
3	Draws conclusion. Bases conclusion on observations comparing three still photos of the fox with the time lapse segment.
2	Draws conclusion. Bases conclusion on observations comparing two still photos of the fox with the time lapse segment.
1	Draws conclusion. Bases conclusion on observations comparing one still photo of the fox with the time lapse segment.
0	Draws no conclusion.

RUBRIC - Sequencing:
TX Essential Element 3 (A) – Classify, order, and sequence data.

POINTS	CHARACTERISTICS
3	Sequences all 4 photos to match actual sequence.
2	Sequences 2 of 4 photos to match actual sequence.
1	Sequences 1 of 4 photos to match actual sequence.
0	Sequences none of 4 photos to match actual sequence.

–Courtesy of AOP Hub/Midwestern State University Assessment Consortium, "Multimedia-Based Assessment Sampler," (Work in progress. Video footage from Optical Data's "Windows on Science," Texas Edition, 1990).

Seamless Web of Instruction and Assessment

If we truly want to ensure that *all* students attain the more complex skills and thinking processes that many feel are important to be successful and productive citizens for the future, we must ensure that we know both what success in these complex areas looks like and what satisfactory progress toward the ultimate goal looks like. Attention to the careful design of assessment forces us to confront these definition issues. Viewed in this manner, good assessment becomes not just an exercise in monitoring set apart from instruction, but the essential ingredient that forces us to be crystal clear about what we want to accomplish with our students, and simultaneously, a tool for helping students attain those goals.

> *Some researchers and educators look forward to the day when there will be a "seamless web" of instruction and assessment in which assessment is no longer a distinct activity but is built into students' regular learning experiences, virtually indistinguishable from instruction. In this vision, students and teachers receive ongoing feedback as needed during the instructional process. Assessment for classroom purposes (as opposed to external monitoring purposes) is a routine, non-threatening process. This vision calls upon students, teachers, and administrators to look at assessment in a different way. Teachers must understand that the primary information will not be quantitative, but will provide a rich portrait of student strengths and weaknesses. In addition, assessment to inform instruction does not have to compare students to one another, which means that student approaches and responses to a problem may look very different and still be correct...[Such assessment] should measure the processes students use as well as the answers they reach. It should measure all of the goals of the curriculum, not just a few. It should address both group activities and individual ones. It should be developed by, or with ample input from, teachers and should include teacher professional observation and judgment. Perhaps most importantly, it should draw upon information from multiple assessment sources, including but certainly not limited to tests.... Finally, it should have a strong self-evaluation component for both students and teachers.* (Kober, 1993, p. 62)

Using Assessment as a Tool for Learning

Introduction

Mention the word "assessment" to most students and teachers and they will immediately respond with the concept of "tests"—the kind used to monitor and track student performance. *TESTS*—the type given to produce grades, select students for special programs, and report to the Board of Education. Wouldn't it be exciting, however, if when we heard the word "assessment," the association was not with traditional testing, but rather with a process that not only (1) provides an opportunity for teachers and students to thoughtfully examine what has been learned, how we know it has been learned, and what needs to happen next to reach a new level of learning, but also (2) is an episode of learning in and of itself?

It *is* possible to use assessments as a significant part of the instructional process. It is possible to have assessment be a tool to demystify the learning process so that teachers and students become equal partners and shareholders. It is also possible for assessment to become a tool for learning in and of itself—students actually learn something from doing the assessment. In this view, one does not teach and then assess, nor is assessment something that one does *to* students—rather, the assessment activity itself *is* an instructional episode.

The following performance assessment, using application of mathematics, is an example of such an assessment.

Performance Assessment: Mathematics

Performance assessment is "assessment based on observation and judgment" (Stiggins, 1994); we observe a behavior or examine a work product and judge its quality. This is frequently done in the classroom by teachers, and is a common occurrence in daily life as well (e.g., the Olympics, job performance reviews, driver's tests). If done properly, performance assessment has the powerful and enabling quality of being a tool for learning.

A major reason that alternative assessment can be a powerful tool for learning relates to the criteria we develop to judge success. When we go through the process of examining student work samples to document varying performance levels, we create a set of criteria that can be used to score the performance. However, the performance criteria can be more than just a mechanism to score or grade performances. They can also provide a very clear communication about what is valued in the work, in other words, what it takes to be successful. Criteria are teachable to students and allow them access to the "secrets" of success. (In fact, students can be full partners in the development of the criteria for success.)

The moral obligation we have to make progress visible to students is given a big boost through thoughtfully written and teachable criteria. Students and teachers begin to share the same vision of success, develop a vocabulary for discussing this vision and, therefore, are able to work toward an identifiable and clear target.

Now, for the mathematics assessment example. Let's begin with analyzing samples of student work. Look at the two student solutions in Figure 3.7 and take a moment to jot down your reactions to them. What did you notice? Are there any obvious strengths? Weaknesses? What features did you respond to? Which solution did you think was the best? Why?

Figure 3.7
Mathematics Applications—Grade 4

Problem: A family of six orders 13 hot dogs. Nine hot dogs have mustard, three have catsup, eight have relish. Explain to the clerk how many hot dogs have no mustard, relish or catsup. You may use drawings in your explanation.

First student response:

DO YOUR WORK HERE

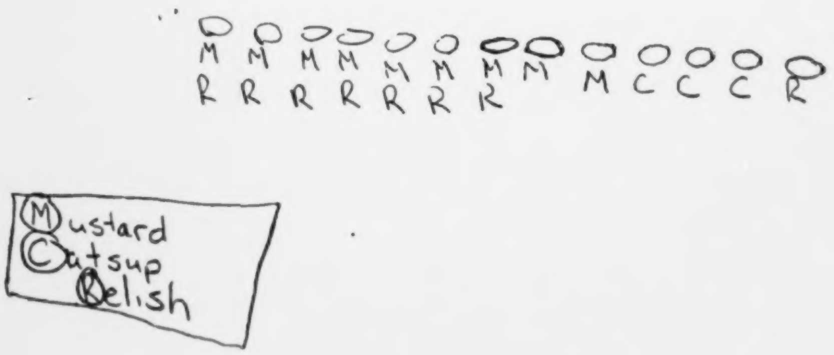
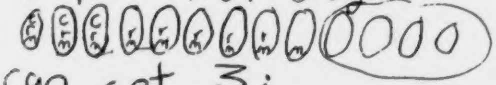


Figure 3.7--continued

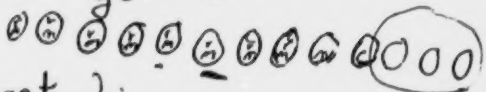
Second student response:

DO YOUR WORK HERE

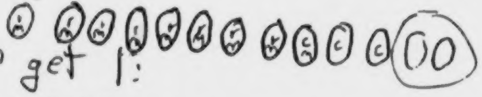
"I would like 13 hot dogs. 9 with mustard 3 with catsup. and 8 with relish." It would be possible to have 4 plain hot dogs or 3 or 2 or 1 or none. This is how you can get 4 plain hot dogs:



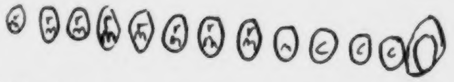
This is how you can get 3:



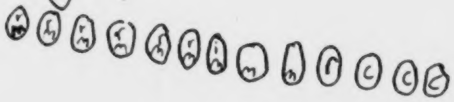
This is how you can get 2:



This is how you can get 1:



You can get 0 this way:



It depends

PS 4

Most teachers examining solutions such as these identify the same sorts of things as features they like and dislike. Typical comments for the first solution would include: *vague, may not have understood the problem fully,* and *had to read a lot into the solution in order to understand what the student did.* Typical comments for the second solution might include: *understood the problem fully, clear solution, communicated well in pictures and words, logical, detailed, organized, insightful, correct answer, explained thinking, used all relevant information, distinguished relevant information from irrelevant information, used data accurately, correct process, on topic, creative, focused, showed process, used pictures correctly, good explanation,* and *went beyond the minimum.* How did your list of comments compare to these?

Comparing solutions to mathematics problems helps us to sort through the complexity of features that make solutions successful. Such exercises can help teachers and students explore what is valued. In fact, the same sets of features re-occur on brainstormed lists from teachers across the country. They form the basis for performance criteria for mathematics applications problems in many places. For example, **Sample 4.2**—*Mathematics Problem Solving* shows a "four-trait" model for assessing such problems. There are four traits because teachers in Oregon felt that their comments tended to fall into four major areas: conceptual understanding, procedural knowledge, problem-solving strategies, and communication. Look at the way teacher comments fit into the four traits in Figure 3.8. Do yours fit the same way? (Other examples of mathematics criteria are **Sample 4.7**—*Performance Assessment in Mathematics (Alberta),* **Sample 5.6**—*Algebra II*, and **Sample 5.7**—*Mathematics Assessment.*)

Figure 3.8
What Teachers of Mathematics Value

Conceptual Understanding

Used all relevant information
Distinguished irrelevant information
Understood the problem
Creative
Used data accurately

Problem-Solving Strategies

Logical
Correct process
Explained thinking
Insightful
Went beyond the minimum
Focused
On-topic

Procedural Knowledge

Used pictures correctly
Correct answer

Communication

Detailed
Organized
Good explanation
Communicated well in pictures and words
Clear solution
Showed process

The four-trait model for analyzing mathematics applications is, thus, basically an attempt to systematically write down *what we already value* so that we can be consistent in our vision of success across students, across assignments, and across time. And, more importantly, so that students can share in the vision of success and not have it constantly shifting in unknown ways between teachers or assignments. (See **Activity 3.1**—*Camping Trip*, **Activity 3.2**—*Sorting Fish*, and **Activity 6.2**—*How Can We Know They're Learning* for activities that involve sorting student work and generating lists of desirable features.)

Value of Criteria to Teachers. Have you ever been nervous grading students' open-ended mathematics problems? Those who have rated or scored mathematics during a direct mathematics assessment using criteria such as the four traits unanimously agree that the very process of assessing teaches a great deal about what makes mathematics successful (as we discovered above). Raters have to internalize a set of criteria so thoroughly that they can be consistent both with themselves over time and with other raters. During the process of scoring hundreds of papers, raters get to see lots of examples of what good and poor mathematics actually looks like and to analyze it systematically to determine why. This allows teachers to go back to the classroom much more confident in commenting on student work and helping students understand what it takes to solve mathematics problems.

If it is true that assessing brings about a shared vocabulary and a deeper sense of what it takes to solve mathematics problems well, then wouldn't students also benefit from the same process? The answer is YES.

Value of Criteria to Students. Students can also learn to look closely and analyze sample work, develop a systematic vocabulary, and put the criteria into their own words so they can develop their own sense of what it means to present a good solution. Consider the following quotes from the area of writing assessment and think of how they might also apply to mathematics problem solving, science process skills, group collaboration skills, and other complex goals we have for students.

> *I want (students) to see evaluation in its best sense—a source to inform, teaching and learning. To that end we develop a vocabulary for commenting on the admirable and problematic aspects of writing....The more we examine samples, the richer and more helpful this language of evaluation becomes.* (Erikson, 1992)

> *Winning points may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game...however often it seems to be about scoring more wins than losses.* (Lucas, 1992)

> *Scales, criteria, and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new material even when they do not have the criteria in front of them.* (Hillocks, 1986)

The goal of assessment should not be only to earn grades. The goal of assessment should be to help students know "how to win again." Clear performance criteria have this effect.

This is assessment that is instructional. Students are learning something in the process of assessing and being assessed. They are learning about mathematics problem solving and applying this knowledge to their own and others' works. They learn what effective mathematics solutions look like and what to do to make their own mathematics skills stronger.

Beyond Writing Assessment

The point here is that good performance criteria are not just useful as a tool for teachers (or districts or states) to use to monitor student ability to write or problem solve, be good collaborators, critical thinkers, etc. Rather, good performance criteria help both teachers and students conceptualize and define standards, especially in hard-to-define areas such as critical thinking, problem solving, collaborative learning, and communication. They provide a vocabulary for discussing features of work or behavior that are subtle and ill-defined.

Good performance criteria help teachers to answer such questions as: *What is expected? What are our standards? What does good performance look like? What do I want to accomplish? What kind of feedback do I give to improve student work next time? They help students to answer similar questions: What is expected? What does good performance look like? How will I know when I'm successful? Why did I win? How can I win again?*

There are two essential features here: (1) having a clear picture in our own heads of the criteria for success, and (2) letting students "in on it."

1. Performance criteria clarify the goals we have for students. They are, in effect, the final and most complete definitions of district and/or teacher goals for students. Notice that it is the **performance criteria** and not the **performance task** that have this function. It is the performance criteria that define what we value in student work, define our targets and goals, and are useful in determining when we get there.

   In many discussions of integrating assessment and instruction, criteria are left out. There is lots of talk about the tasks that should be given to students and how one watches performance on these tasks to monitor where students are. But the question is: *What, exactly, do you look for?* in the performance in order to determine progress, know when students are ready to move on to the next topic, or determine when mastery has occurred? The *What do you look for?* are criteria. Overt and public criteria, arrived at through discussion and collaborative learning and accessible to teachers, students, and parents bring judgmental assessment from the realm of individualistic impressions to the realm of systematic and generalizable documentation of learning. They also result in assessment to promote learning.

   But, not all performance criteria are created equally—they may not all be able to have this effect. There may be certain design features that make them more or less useful as instructional tools. These features are outlined in Chapters 4 and 5 and **Activity 4.3**—*Performance Criteria—Keys to Success*.

2. Students must be involved in assessing. Students learn to the extent that they are the ones doing the judging.

   By encouraging students to engage in self- and peer evaluation teachers empower students to take control of their own learning. When students help determine the criteria for assessment they can make reasonable decisions about the quality of their own work. By engaging students in self-assessments, students learn they are ultimately responsible for their own learning. It is in this area that portfolios are so powerful. (Tierney, et al. 1991, p.59)

   As a matter of fact, one perspective is that engaging students in conscious development and systematic application of criteria is a major attribute of critical thinking. Therefore, the process of engaging students in assessment actually helps to attain some of the thinking goals we have for them.

> *To be successful, students must learn how to evaluate rather than how to be evaluated....Evaluation is routine, something to be done at every step of the way as they examine decisions, observations, data, ideas and outcomes....[When] evaluation is a key process, and one over which students have considerable control, they are truly critical thinkers.* (Penick, 1989)

Relationship Between Classroom Performance Assessment and Large-Scale Assessment

So far, it may seem that we are only discussing classroom performance assessment. But, in reality, the whole enterprise is much more powerful if large-scale and classroom assessment are considered together as a unit. Consider the following points:

1. What if performance criteria for large-scale assessment (which have to be developed anyway) were to be developed by teachers with classroom uses in mind? In order to develop performance criteria, teachers have to engage in the discussions and do the research necessary to develop a common vision of what it means for students to be successful. The effort that goes into large-scale development is also, therefore, (a) staff development and (b) development of instructional materials.

2. What if the performance criteria were uniform across teachers, so that we have confidence that regardless of which teacher is observing a student performance, the judgment would be the same? Then assessment credibility improves—and students get the same message across teachers and grades. Large-scale assessment development can take performance assessment out of the realm of individual interpretation and can promote essential dialogue about what, in fact, constitutes our vision of success for students.

3. The way we do large-scale assessment sends messages to teachers on (a) what to teach and (b) how to teach it. If we design large-scale assessment with the instructional end user in mind, we have much more opportunity to influence instruction in the way we intend.

References

Costa, A., & Kallick, B. (in press). Reassessing assessment: Seven issues facing renaissance schools. In R. Blum & J. Arter (Eds.). *Now showing: A resource guide for student performance assessment in the context of school restructuring.* Alexandria, VA: Association for Supervision and Curriculum Development.

Erikson, M. (1992). Developing student confidence to evaluate writing. *Quarterly of the National Writing Project & The Center for the Study of Writing and Literacy, 14*, pp. 7-9.

Evans, C. (February 1993). When teachers look at student work. *Educational Leadership 50*(5), pp. 71-72.

Fennema, E., Carpenter, T., & Peterson, P. (1989). Learning mathematics with understanding: Cognitively guided instruction. *Advances in Research on Teaching, 1*, pp. 195-225. Greenwich, CT: JAI Press Inc.

Gong, B., Venezky, R., & Mioduser, D. (1992). Instructional assessments: Lever for systemic change in science education classrooms. *Journal of Science Education and Technology, 1*, pp. 157-176.

Herman, J., Aschbacher, P., & Winters, L. (1992). *A practical guide to alternative assessment.* Los Angeles: The Regents of the University of California.

Hillocks, G., Jr. (1986). *Research on written composition.* Urbana, IL: National Council of Teachers of English.

Kober, N. (1993). *EdTalk. What we know about science teaching and learning.* Washington, DC: Council for Educational Development and Research.

Land, M. & Peters, T. (in progress). *Multimedia-based assessment sampler.* Wichita Falls, TX: Midwestern State University, Division of Education.

Lucas, C. (1992). Introduction: Writing portfolios—changes and challenges. In K. Yancey (Ed.). *Portfolios in the writing classroom*, pp. 1-11. Reston, VA: National Council of Teachers of Mathematics.

Measuring up: Prototypes for mathematics assessment. (1993). Washington, DC: National Academy Press.

Murphy, S., & Smith, M. (1990, Summer). Talking about portfolios. *The Quarterly of the National Writing Project.* Berkeley, CA: University of California, Center for the Study of Writing.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

Oregon Department of Education. (1991). *Mathematics problem solving scoring guide.* Salem, OR: Author.

Penick, J. (1989). Issues-based education—A need and rationale. *Education and Urban Society, 22*(1), pp. 3-8.

Perrone, V. (Ed.). (1991). *Expanding student assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Regional Educational Laboratory Network Program. (1994) *Improving science and mathematics education—a database and catalog of alternative assessments, second edition.* Portland, OR: Northwest Regional Educational Laboratory.

Shavelson, R., & Baxter, G. (1992). What we've learned about assessing hands-on science. *Educational Leadership, 49*(8), pp. 20-25.

Spector, B. (1993, January). Order out of chaos: Restructuring schooling to reflect society's paradigm shift. *School Science and Mathematics, 93*(1), pp. 9-19.

Stayter, F. & Johnston, P. (1990). Evaluating the teaching and learning of literacy. In T. Shanahan (Ed.), *Reading and writing together: New perspectives for the classroom.* Norwood, MA: Christopher-Gordon Publishers, Inc.

Stiggins, R. (1994). *Student centered classroom assessment.* Columbus, OH: Macmillan.

Sugarman, J. (1989, November 27). Portfolios useful assessment tool. *Education USA, 32*, p. 101.

Tierney, R., Carter M., & Desai, L. (1991). *Portfolio assessment in the reading-writing classroom.* Norwood, MA: Christopher Gordon, Publishers, Inc.

U.S. Department of Education. (1991). *America 2000: An education strategy.* Washington, DC: Author.

ACTIVITIES

Activity 3.1

Camping Trip: A Math Problem-Solving Task

Purposes:

1. To articulate valued characteristics of student math problem solving
2. To experience the usefulness, both as staff development and as an instructional tool for students, of the process of developing generalized performance criteria
3. To experience where criteria come from by "reconstructing" a scoring guide
4. To see that alternative assessment is "do-able"

Uses:

1. In Chapters 3, 4 or 5 to illustrate how the development of performance criteria can also be an integral part of professional development for teachers or an instructional activity for students
2. In Chapters 3, 4 or 5 to illustrate how to develop generalized performance criteria

Rationale:

Well-designed performance criteria are not just something others have pulled out of a hat. They actually represent *what we value* in student work. We currently judge the quality of student responses all the time, but we typically don't formally write down the basis on which we judge. The advantage of having overt criteria that are written down is that we, as teachers, can examine the soundness and clarity of our vision of student success and be more consistent in our subjective evaluation of student work across different assignments, across students, across time, and between teachers. Also, there is an advantage for students in clearly articulating criteria. It becomes very clear for them what is expected, and they can use the same criteria to evaluate their own work. (In fact, there is an entire instructional technology built around student development and use of performance criteria.)

This exercise illustrates the power of the *process* of developing performance criteria. The previous sections of Chapter 3 present this argument in more detail. The specific exercise involves the development of performance criteria for mathematical problem solving.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Set of student answers to problem-solving exercises in math; 8 to 10 responses representing different levels of sophistication are optimal; one such set is included in this training package (**A3.1,H1**); users are encouraged to generate their own set of problem solutions
- Oregon's four-trait analytical scoring guide for mathematical problem solving (**Sample 4.2**); one copy per participant; (other generalized math problem-solving scoring guides could also be used, for example, **Sample 4.7**—*Performance Assessment in Math [Alberta]* or **Sample 5.7**—*Mathematics Assessment*)
- Chapter content to use as a handout: Using Assessment as a Tool for Learning
- Optional: *Build Your Own Performance Criteria* (**A3.1,H2**)

Time Required:

90 minutes

Facilitator's Notes:

1. Explain the purpose of the exercise. Use the rationale presented above.
2. Have participants work in groups of two or three. Give one set of student problem solutions to each group (**A3.1,H1** or sets of locally generated student work).
3. Participants should put the student solutions into three piles based on the "quality" of the solution. Describe this task in ways that cue what to do. For example, you can frame the task in terms of grading if it will help participants understand what to do: *Put papers into stacks by what grade you would give to them.* Or, the three stacks could represent levels of sophistication.

   As participants are sorting the solutions, they should write down the basis on which they sorted: *You thought the papers in this stack were better than the papers in this other stack. Why? What features in the work did you like? Not like? What makes the work in this stack different from the work in that stack?*

   The goal is not to score all the papers but to articulate reasons. Therefore, they only need to continue sorting and discussing until all their reasons are listed. Remind each group of this as they work.

4. Put up a clean overhead transparency with three columns labeled "high,""medium," and "low" or "beginning,""developing," and "fluent." (Use whatever titles will

appeal to the group.) Have groups tell you the placement of each paper. If there is a disagreement about placement, write the paper number under a "?" heading.

There is not a "right" answer for this sorting. The goal is not to sort "correctly," but to have participants articulate what they value, come to consensus, and be able to justify their judgments using consistent terminology.

5. List all the criteria participants had for their sorting.

6. Try to resolve ratings of the "?" papers. Have the various groups describe their rationales for placement. Add to the criteria list as new or expanded points emerge. The goal is to have a rich and elaborated list of criteria that cover many features of what is valued.

   Here's a list generated at a recent training: complete, elegant, understandable, efficient, not confusing, correct calculations, right answer, checked answer, understood the concepts, used mathematics correctly, used estimation correctly, tried several different strategies, used a novel approach, made a connection to another problem, made good choices, knew which information was needed, didn't guess, systematic, used alternative representations of the problem, not a rote application of an algorithm, organized, coherent, sufficiently detailed, clear, comprehensible, didn't have to guess what the student meant.

7. Take the list of criteria generated by the team and see where they fit in the four-trait model. The point here is that they have just recreated the four-trait model. (Of course, it is perfectly acceptable to modify the four-trait model as well. In fact, the desire to modify the wording or traits indicates that participants feel that the model does not quite describe what *they* are seeing in student work. This is good! We need to make criteria our own, so that they communicate to *us* what works in student products.)

   (Alternative to step 7: Finish the steps of developing performance criteria as shown on **A3.1,H2**.)

8. Take a new solution to the same problem or a new solution to a new problem and try to score it using the four traits. On each trait have participants determine whether the paper is "stronger than weak" or "weaker than strong." Then have them determine how weak or strong. Vote on scores. Participants should justify their judgments using phrases from the scoring guide that *describe* what they are seeing.

   Go through all four traits on enough problems that participants are reaching agreement.

9. Ask the participants to discuss what they learned from this activity. Do they think students would benefit from doing this? What might students learn? How else might this be used with students?

Sample 1

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

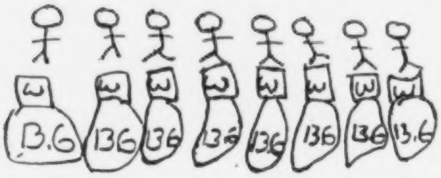
Explain your answer.

If 12.5 liters are needed for 5 people for 1 day well then 45 liters will be needed for three days for 8 people because for 1 person it was 1.5 which all together equaled 12.5 for 1 day, for 8 people to go on a 1 day trip you'd have ~~to~~ bring 15 liters.

Sample 2

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer.



$$\begin{array}{r} \overset{1}{1}2.5 \\ 3 \\ +\ 8 \\ \hline 13.2 \end{array}$$

liters of water
13.6 l for 8 people to bring camping for 3 days

A3.1,H1

BEST COPY AVAILABLE

Sample 3

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer. 60 Liters

I came up with that by getting how much 1 person needed Then how much 8 people needed that was 20 and then I multaplied that by three for three days and came up with sixty

Sample 4

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer. 20. liters for 8 people one day

2.5 for each person

$$\begin{array}{r} 12.5 \\ \div\ 5 \\ \hline 2.5 \end{array} \qquad \begin{array}{r} 12.5 \\ +\ 2.5 \\ \hline 15.0 \\ +\ 2.5 \\ \hline 17.5 \\ +\ 2.5 \\ \hline 20.0 \end{array}$$

A3.1,H1

Sample 5

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer.

2.5 liters/person
5 people) 12.5 liters

2.5 l/person
x 8 people
20 liters/day

20 liters/day
x 3 days
60 liters in all

I divided 12.5 liters ÷ 5 people = 2.5 liters/person. I did that so that I could take 2.5 liters x 8 people = 20 liters/day. Now I need to multiply 20 liters/day x 3 days = 60 liters to last the whole camping trip. 60 Liters in all.

Sample 6

1
12.5 litters
x 2
4
25.0 litters
x 8 people
200.0 litters
x 3 days
600.0 litters are needed for 8 people for 3 days

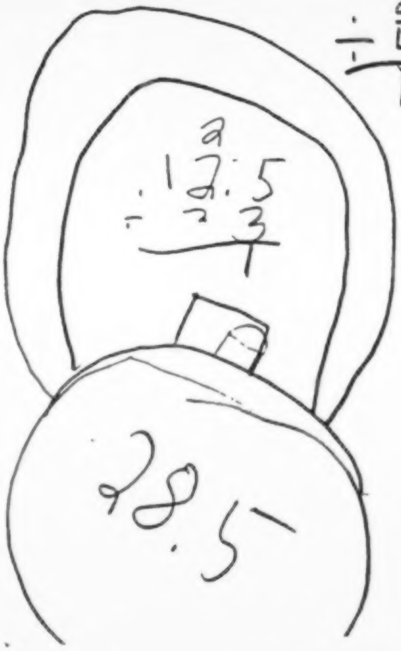
A3.1,H1

Sample 7

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer.

12.5
÷5
4



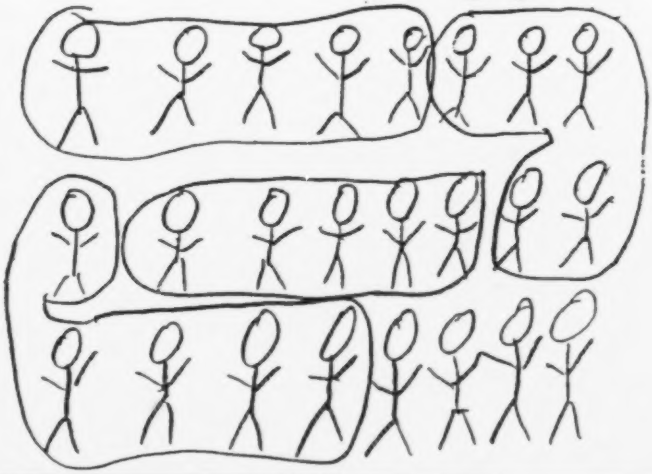
12.5
+2
24.7

28

24.5
+4
38.5

Sample 8

50.



A3.1,H1

Permission for use in training given by
Lewiston School District (Idaho)

Sample 9

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer.

2.5 lieters of water apiece

5 people) 12.5 lieters of water for a group of 5 people.

2.5 lieters each of water
x 8 people
20 lieters of water needs to be token at 2.5 lieters each

A3.1,H1

BEST COPY AVAILABLE

Sample 10

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer.

12.5
× 2 more people
25
× 2 more days
50

info
much 12.5
8 people 3 days

First e gathered some importan info. 5 peopl for one day was 12.5 e add two people by timing to equel 25. then e times it by two again for the day e new that one day was all ready couned 50 liters

A3.1,H1

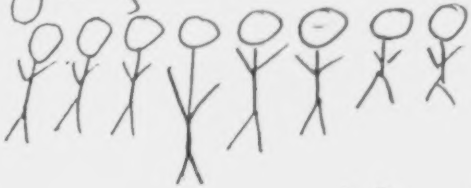
Permission for use in training given by Lewiston School District (Idaho)

Sample 11

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer.

8 = 3



12.5
÷ 5
2.5
× 4
10
+50
60

Sample 12

14. A group of 8 people are all going camping for three days and need to carry their own water. They read in a guide book that 12.5 liters are needed for a party of 5 people for 1 day. Based on the guide book, what is the minimum amount of water the 8 people should carry all together?

Explain your answer. They can bring enough water for 5 people for 6 days which 75 liters which should be enough

A3.1,H1

BUILD YOUR OWN PERFORMANCE CRITERIA

Sample Topic: Self-Reflection

Here's how you develop general performance criteria. I will illustrate the steps using student self-reflection as an example. Remember, we are *not* developing these performance criteria so we can "grade" student self reflections. We are doing it as an exercise in clear thinking about one of the goals we have for students.

Step 1: Gather samples of student performance

Gather samples of student performance that you feel illustrate the skill or behavior in question. In our case we are trying to assess student self-reflection skills, so we might collect things that students have written in answer to questions like: "Select a piece of work that you feel illustrates your best effort. Why did you select this piece? What does it show about you as a ______?" (The blank would be filled by "writer," "math problem solver," "lifelong learner," or whatever skill you are focusing on.) Other questions that might elicit student self-reflection might be: "How have you changed this year in your ability to _________?" or "What do you currently understand about ______? What do you still need to understand?"

When you gather samples, it is important to represent a range of skill levels.

Step 2: Brainstorm a list of attributes

Place the samples of student work into three piles: strong, middle, and weak. For our example we might have three piles that illustrate levels of sophistication with respect to self-reflection. As you sort the student work, write down the reasons you have for placing pieces in the various stacks. In other words, if you place a piece in the "sophisticated" pile, why? What are you seeing in the work that makes it "sophisticated?" Write down all evaluative comments; they reveal criteria.

Keep reading student work and placing it into piles until you are not adding anything new to your list of attributes. Try to create as large and diverse a list as possible. Recent lists have included such things as: detailed, many things covered, insightful, self-revelation, examples provided, motivation, sets goals for the future, looks at more than one thing, considers content as well as process, takes risks, accuracy, discussion related to criteria, growth supported with examples,

A3.1,H2

organized well, sincerity, honesty, comparisons over time, shared feelings, looks at strengths and weaknesses, depth of analysis, good reasons and explanations, revealing, voice, easy to read, looks at skill improvement, ownership, personal reaction, specificity, looking ahead, thorough, synthesizes ideas, readable, neat,... While doing this initial brainstorming for a real, educational application, it is a good idea to consult the professional literature so that you can be sure to be as up to date as possible on the characteristic of interest. For example, if you are developing criteria for critical thinking, it is a good idea to read up on critical thinking. This will help you articulate the important dimensions of performance.

Step 3: Cluster

Cluster similar attributes together and begin to form "traits" or "dimensions" of self-reflection. Try to limit the number of traits (three to six is good). For example, the above attributes *might* be clustered into the traits of: **SKILL ANALYSIS, SINCERITY, GOAL SETTING,** and **PRESENTATION.** Make sure that your "traits" cover everything. For example, **SKILL ANALYSIS** might cover the attributes of: detailed, many things covered, insightful, self-revelation, examples provided, looks at more than one thing, considers content as well as process, accuracy, discussion related to criteria, growth supported with examples, looks at strengths and weaknesses, depth of analysis, good reasons and explanations, revealing, specificity, thorough, and synthesizes ideas. **SINCERITY** might cover: motivation, takes risks, sincerity, honesty, shared feelings, voice, and ownership. **GOAL SETTING** might cover: looking ahead and sets goals for the future. **PRESENTATION** might include: readable, neat, organized well, and easy to read.

Step 4: Write a value-neutral definition of each trait

Write a definition of each trait. For example: *Sincerity is the degree of ownership, effort, and honesty in the self-reflection. It is the degree to which the student makes a sincere effort to self-analyze.* These definitions should be "value neutral"—they describe what the trait is about, not what good performance looks like. Here is an example of the definition above converted to a statement that is not value-neutral: *Sincerity is taking ownership for work, really trying hard, and being very honest; the student has made a sincere effort to self-reflect.*

A3.1,H2

Step 5: Describe strong, middle, and weak performance on each trait

Generate descriptions of strong, middle and weak performance on each trait. Use adjectives and descriptive phrases. Remember, performance criteria describe performance so that you can match up what you are seeing in student work to the descriptions provided.

Example: **SKILL ANALYSIS**

Strong (5)	**Middle (3)**	**Weak (1)**
Specific, clear examples	I like it because...	Vague
Unique	Shows, then hides	Simple restatements
Thoughtful	Beginning of ownership	Mechanical
Reflective	Compare/contrast	Focuses only on surface features
Interesting	What I know	Obvious
Examines process skills	Two-dimensional	Rote response/Same old
Purposeful	Some specifics	No examples
Lots of examples	Describes performance	Purposeless
Tells the whole story		I like it/I don't like it
Three dimensional		One-dimensional/superficial

Step 6: Benchmarks

Find samples of student work that are good examples of strong-, weak- and mid range. Be sure to have several sets representing typical ranges of performance at each level of early elementary, middle, and high school. These can be used to illustrate to students what to do.

Step 7: Make it better

Criteria evolve with use. Try them out. You'll probably find some parts of the criteria that work fine and some that don't. Add and modify descriptions so that they communicate better. Choose better benchmark and sample papers. Revise traits if you need to. Let students help—this is a tool for learning.

Judy Arter, Vicki Spandel, Ruth Culham, Rick Stiggins
March 1994

A3.1,H2

Activity 3.2

Sorting Fish: A Science Task

Purposes:

1. To sort student papers into categories based on criteria participants value or already look for in student work
2. To develop a scoring guide for the "sorting fish" task
3. To discuss:
   a. The characteristics of quality student work
   b. The differences between developing (or modifying) criteria and/or rubrics *after* the task has been administered to students, rather than *before*
   c. The instructional strategy implications of engaging students in developing generalized performance criteria and rubrics

Uses:

1. A science alternative to **Activity 3.1**—*Camping Trip*
2. In Chapters 3, 4, or 5 to illustrate how the development of performance criteria can also be an integral part of professional development for teachers or an instructional activity for students

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overheads A3.2,O1**—*Sorting Fish,* **A3.2,O2**—*Scoring Criteria and Rubric,* **A3.2,O3**—*Focus Questions*
- **Handout A3.2,H1**—*For Kids, This Is No Fish Story*
- Set of 13 student responses in **A3.2,H2**—*Sorting Fish,* laminated if desired

Time Required:

60-90 minutes

Facilitator's Notes:

1. Explain the purpose of the activity: *to examine some student papers generated in response to an assessment task administered to a group of fifth grade students and to categorize them by quality.*

2. Have participants work in groups of two to four. Use **Overhead A3.2,O1**—*Sorting Fish* to present participants with the assessment task. Then give each group a set of the 13 *Sorting Fish* **(A3.2,H2)** student sample pages. Instruct groups to: *read each student response and sort them into three categories based on their perceived quality, e.g., grade they might receive (-, √, +), (low, medium, high), etc.*

   You may need to walk around the room and prod groups into sorting; some will resist the three-pile limitation; others may be unable to agree on a basis for sorting. (15-20 minutes)

3. As groups complete the sorting process, ask them to note their rationale for placing papers in each category. (Note: These will form **criteria** for sorting, although participants may not use that label.) *What makes stack 1 different from stack 2 and stack 2 different from stack 3?*

4. Prepare a transparency with three blank columns. Label the columns "Low," "Medium" and "High" (or any other labels that appeal to the participants).

5. Once sorting is completed, have groups report on the category for each student paper, while you record their decisions on the transparency. Note any discrepancies as to how a sample is categorized. (10-20 minutes) *Into which stack did you place each of the student papers?*

   OR, an alternative is to have groups report on their rationale for placing papers in each group, while you record the reasons on the transparency. Use these to begin step 6.

6. Prepare a separate transparency, with two columns, one labeled "General" and the other "Task Specific." Explain the difference between these two columns: *The "General" column will include criteria that could be applied to any task, while the "Task Specific" column will include criteria specific to this particular task of "Sorting Fish."*

   As participants call out their reasons/criteria for placing items in categories, ask the large group to help determine into which columns the criteria should be placed:

Would that reason be specific to this "Sorting Fish" *task or could it apply to many other tasks?*

List all the criteria participants had for their sorting. (5 minutes)

At a recent training, the following criteria were suggested:

a. Task specific: technical vocabulary specific to this task, fish named, comparison with initial fish

b. General: coherence, specificity, multiple comparisons, writing ability, number of identification points

You may also wish to ask whether or not general and task-specific criteria should be mixed in developing scoring rubrics, e.g., advantages and disadvantages of having two separate rubrics: a general set and a task-specific set. Other discussion questions: *How have broad categories been included, such as factual recall, process skills, writing skills, or higher-order thinking skills? Are such inclusions important criteria?*

7. If there were discrepancies in how papers should be sorted or in the rationale for placing papers in certain categories, ask participant groups to try to resolve these based on the criteria listed. (5 minutes)

8. At this point, you may wish to share the following background information with participants: *In the fall of 1993, a group of fifth grade science students from Washington/Jackson Math Science Center, Wichita Falls, Texas, assisted in the scoring of live fish taken from a pond drained by the city's Parks and Wildlife Department* (see attached article, **A3.2,H1***)*. *Students completed the* Sorting Fish *assessment (designed primarily to assess classification skills) following the hands-on fish classification, and follow-up classroom activities involved the use of fish classification keys.*

9. Ask groups to use their criteria and reasons for placing papers in each of the categories to develop a three-, four-, or five-point scoring guide for the *Sorting Fish* task. (10 minutes) *You will now use your criteria to develop a scoring guide for the* Sorting Fish *task. Select one of the criteria. Beginning with "1" as the lowest score, write a description of the least acceptable form in a student response to get a numerical score. For example, if you decided to use the task-specific criterion of being able to correctly name the fish, you might assign the following description a value of 1: "The student correctly names at least one kind of fish." Two or three correctly named fish might be assigned a value of 2, and four or five fish, a value of 3. While the descriptions may be short (one to two sentences), be certain not to sacrifice clarity for brevity. As another example, consider the general criterion of "coherence." What would a score of 1 look like? A score of 3? A score of 5?*

10. Ask several groups to read some of their sentences. Identify these statements as **rubrics**.

11. Display **Overhead A3.2,O2**—*Scoring Criteria and Rubrics*, presenting it as a rubric modified from one developed by Cheryl Wilson, the supervising teacher of the *Sorting Fish* project and task, *prior* to administering the assessment. Ask participants what information Cheryl had that the participants might not have had. (5 minutes)

12. Use **Overhead A3.2,O3**—*Focus Questions* to guide participants in small group discussions of the three focus questions. (5-10 minutes)

    a. **What represents quality student work?** Points that need to be made:

    - Quality work depends not only on getting the right answer or answers, but on being able to make appropriate connections, to apply knowledge in other contexts, and to create new knowledge based on insight.

    - There needs to be a consensus between teacher and students about what constitutes quality. Thus, the game of "guessing what the teacher wants" is minimized while students' confidence in their thinking and reasoning abilities is enhanced.

    b. **What are differences between developing or modifying criteria and rubrics *after* the task is administered to students, rather than *before*?** Points that need to be made:

    - Sometimes, by setting criteria before the task is administered, the chance of bias in scoring is decreased. With established benchmarks, the extent to which students meet expectations can easily be determined. This is especially true for task-specific scoring.

    - Even though the common practice is to develop or modify criteria *after* the task is administered, this practice encourages comparison of student responses to each other rather than to a set standard. Thus, the score might not reflect student ability to perform the task adequately, but could instead reflect the quality of student responses in relation to the responses of other students. Ultimately, we want criteria ahead of time. However, during the development of criteria, we frequently use samples of student work to help us sharpen our thinking about what's valued. Therefore, during development, criteria can be modified in response to an examination of student work.

    c. **What are the implications of engaging students in developing generalized performance criteria and rubrics, as you have just done?** Points that need to be made:

- By engaging students in the process of developing criteria for a specific task, the chance of misunderstanding regarding the basis on which students are assigned a score for the task will be significantly reduced. That is, they will know up front what is expected of them in a response.

- As an instructional strategy, this process can effectively be used both to guide students in increasing their ability to conduct self-assessments and to encourage them to question and seek out criteria for other assignments.

13. Summarize the main points of the activity by having groups report on the results of their discussions. (5-10 minutes)

Courtesy of *Wichita Falls Times Record News*, Wichita Falls, Texas, September 22, 1993

Photo/Becky Chaney

Fishing for facts

Barbara Maruscsak, fifth-grade teacher at Washington-Jackson Math-Science Center, helps Nick Stephens and T.J. Chancellor weight fish from the lake in Williams Park. The kids are helping the Parks and Wildlife Department sort and tag the fish. Story, 6B.

For kids, this is no fish story

By Christy Ward
Staff Writer

The fish from Williams Park Pond will soon have a new home.

As part of a class project, the fifth graders from the Washington-Jackson Math-Science Center helped the Texas Parks and Wildlife Department drain the pond Tuesday and sort the fish to be taken to different ponds in the city.

"They're going to drain it because it has a leak in it," said Kristina Romoser, 10.

The fish squirmed through the students' hands as they helped decide their different species.

"We're learning about what kind of fish they are," said Maralyn Newman, 11.

The project is part of the Rally Round the River project for the fifth-grade curriculum at the math-science center. Throughout the year they will learn about the river and use the informtion gathered for science, math and language-arts projects as part of the team effort of learning, said Cheryl Wilson, science teacher.

First, the students classified the fish for science so they could learn about the different kinds of markings, said Adrian Wainscott, 10.

The next step was for each child to use the measuring tapes that they made of straws to find the length and a scale to measure the weight.

"The fish will help us learn about measuring," said Keandress Brown, 10.

The fifth graders will continue to do similar projects with their mentor, the Texas Parks and Wildlife Department, which presented a slide show last week to the classes about habitat changes for fish, Wilson said.

"One of their goals is for children to get more involved in fishing," she said.

This project helped the students use the information they learn in the classroom and apply it to real life situations, said Karin Evans, math teacher.

"They are especially getting a good concept of estimating," said Barbara Maruscsak, math teacher. "It is one of the few times they actually get to apply metrics."

The wildlife department will also help stock the ponds at the school by having the kids grow the fish from the time they are eggs on up, Wilson said.

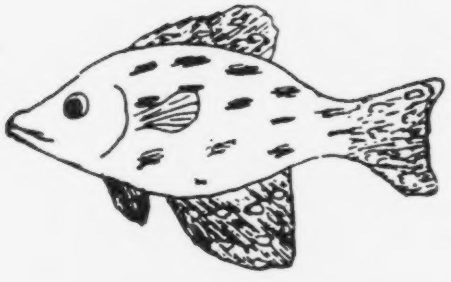
"We're learning how fishy of a job it is," said Stephanie Husak,

BEST COPY AVAILABLE

A3.2,H1

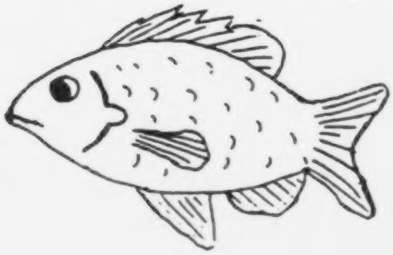
Sorting Fish

Sample 1

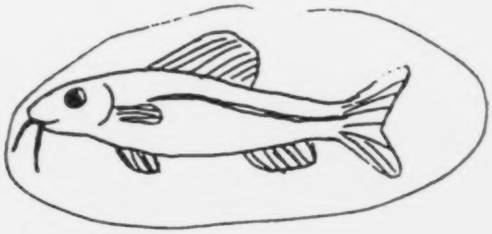


Which of the fish below is not like the above fish?

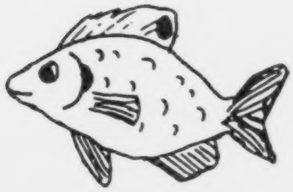
A.



B.



C.



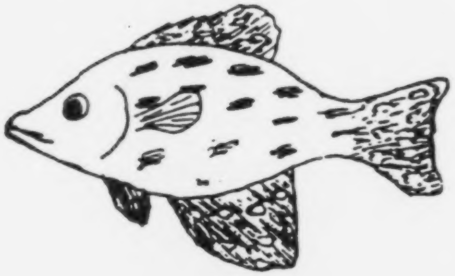
Now explain why.

Fish "B" the Catfish it is the longest fish there and it is the only fish with wiskers. It is the skinniest one out of all of the other fish. Its dorsal fin go's up higher and it's dorsal fin doesn't have spikes. He has two lines running down his side.

A3.2,H2

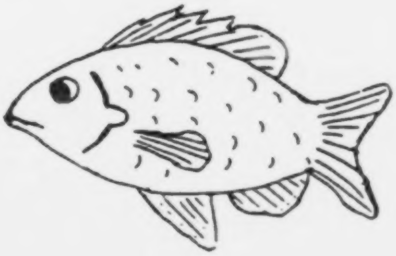
Sorting Fish

Sample 2

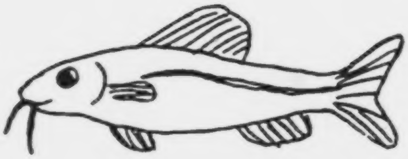


Which of the fish below is not like the above fish?

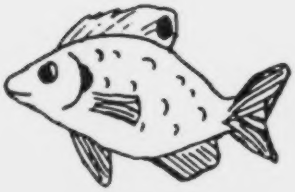
A.



[✓ B.]



C.

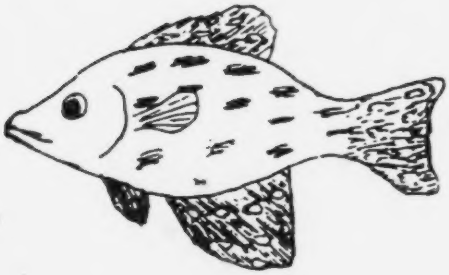


Now explain why.

One is a Croppie and
One is a Cat Fish
One has wiskers and
One dosten.
One has lines on his
tales and one donez't
One has ≈ on it on its
Body and one had ▬
on it's body

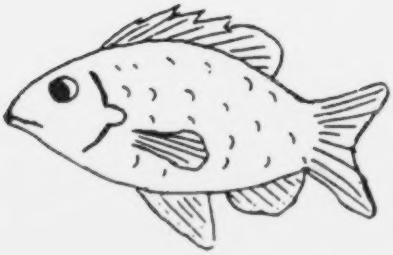
Sorting Fish

Sample 3

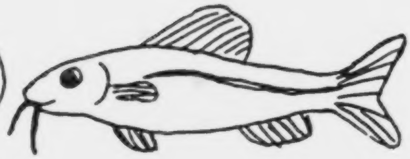


Which of the fish below is not like the above fish?

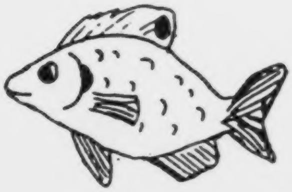
A.



[✓ B.]



C.



Now explain why.

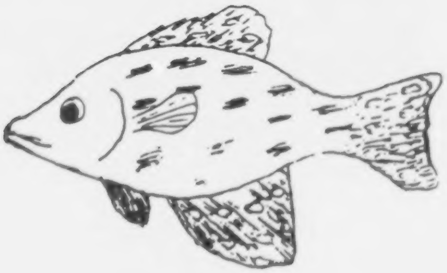
Catfish B

Smaller - Skinny

It does not look nothing like the other fish, his mouth is not pointed. No desine or his top fin. His first bottom fin by his mouth is not kinda. One spine sticks his spine

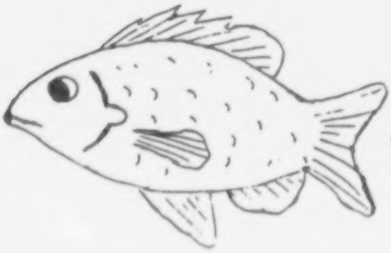
Sorting Fish

Sample 4



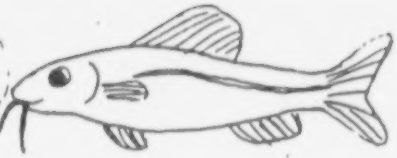
Which of the fish below is not like the above fish?

A.



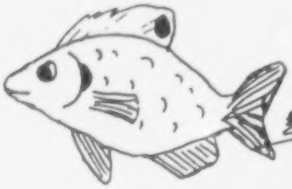
Shad

[✓ B.]



~~Sap~~ Channel catfish

C.



bluetail

Now explain why.

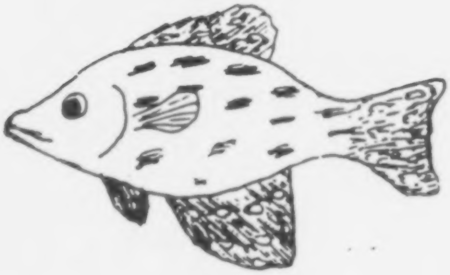
The catfish has whiskers
The croppy does'nt
The catfish has a line
The croppy does'nt
The catfish has fork tail
the croppy does'nt
The catfish isn't fat
The croppy is

165

A3.2,H2

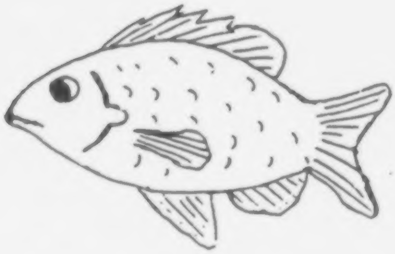
Sorting Fish

Sample 5

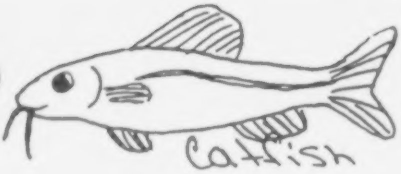


Which of the fish below is not like the above fish?

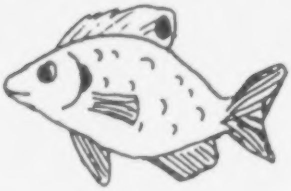
A.



[✓ B]



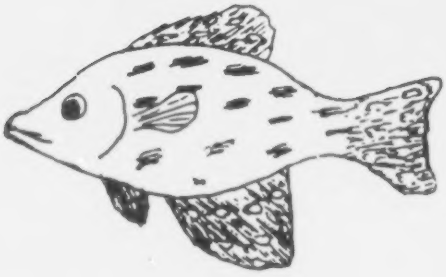
C.



Now explain why.

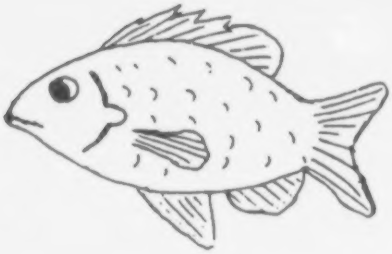
It has whiskers under it's month, it's tail is shapet like a V. It has a line From its fin to his back fin. Catfish don't have Spots. Catfish are also pretty long. The Catfish weighs more than the other fish. It also has a back bone.

Sorting Fish
Sample 6

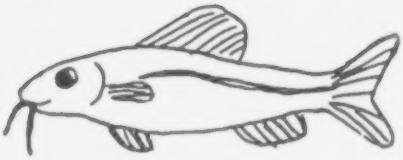


Which of the fish below is not like the above fish?

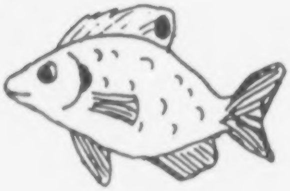
A.



[✓ B.]



C.



Now explain why.

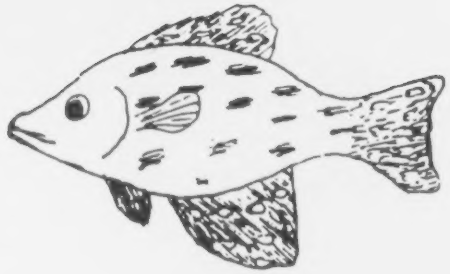
Beacouse it has a
set of whiskers
his fins arnt spikey
difrent desighn
his lips arn't
one over a nother
he dont have spots
he has littler fins
he desighn is
lines he is
skiny and
is real long
and not to big
he only has
one line on
his body
one fin is
not realy bigger
than the othe
finns. The
other on dont
have two
whiskers his
Mouth is a lot
liter than
the other fish

BEST COPY AVAILABLE

A3.2,H2

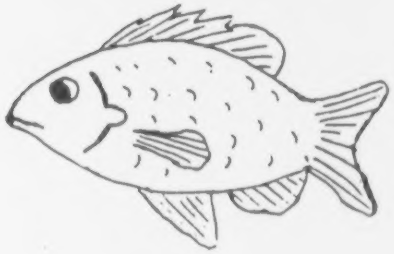
Sorting Fish

Sample 7

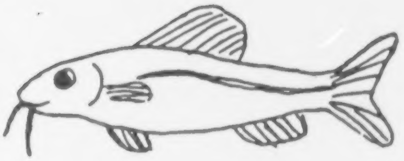


Which of the fish below is not like the above fish?

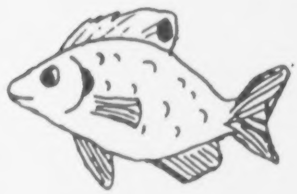
A.



[✓ B.]



C.

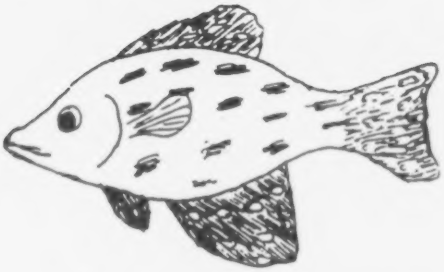


Now explain why.

The cat fish has wiskers and none of the others do. The cat fish is long and slender and the others are short and fat. The catfish has a long line on its back the others have gills. The cat fish's tail is sharp looking the cropy's is rounded on the ends. The bluegill has a spot on its dorsal fin and the catfish does not. Like I said at the beganing the catfish has a long line on its back the cropy has like little spots.

A3.2,H2

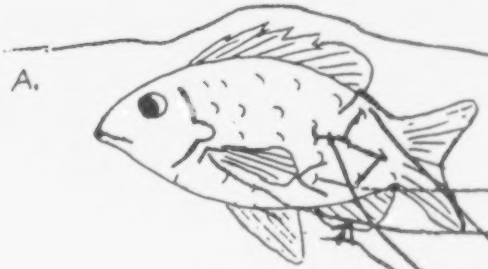
Sorting Fish
Sample 8



Which of the fish below is not like the above fish?

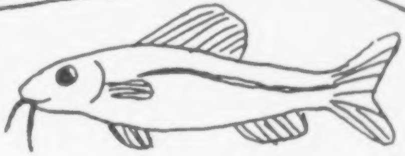
Now explain why.

A.

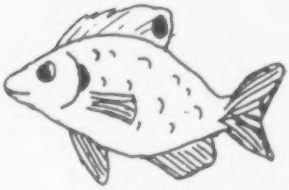


I think it does not ~~looks like~~ from the : is lower and the has a lot in it and the took pointer and the loops too and the is shorter

B.



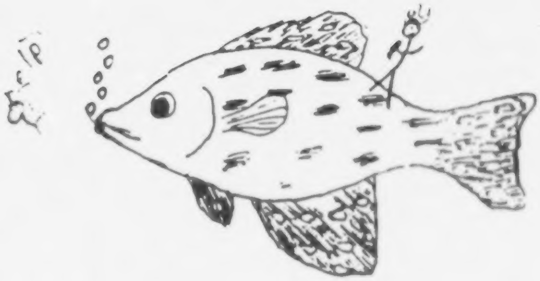
C.



A3.2,H2

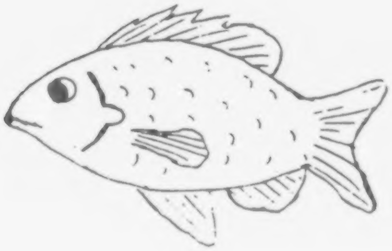
Sorting Fish

Sample 9

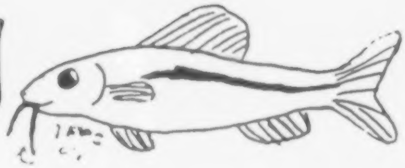


Which of the fish below is not like the above fish?

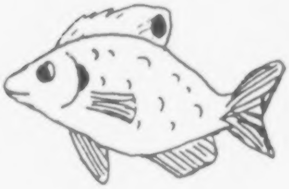
A.



[✓ B.]



C.

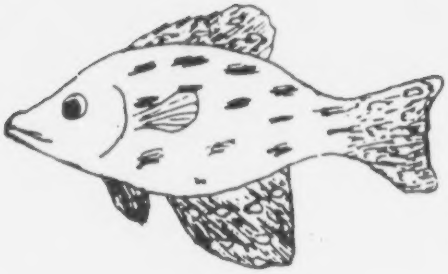


Now explain why.

Becase the one at the veary top has spots so doses A and C But B don't it has wiskers and the crape dosent. The crope has spikes on his gills and the cat fish dosent, and the crope is Bigger than the cat fish and the cat fish has a line and the crop dosen't

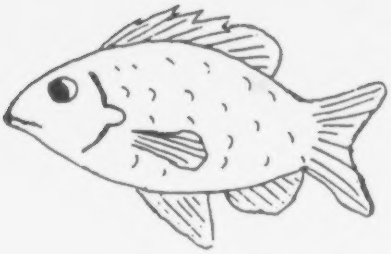
Sorting Fish

Sample 10

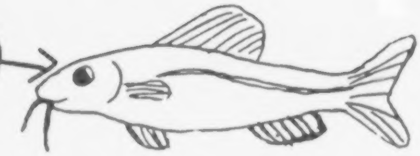


Which of the fish below is not like the above fish?

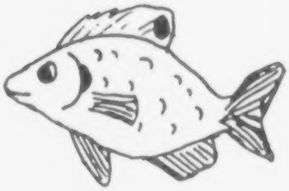
A.



[✓ B.]



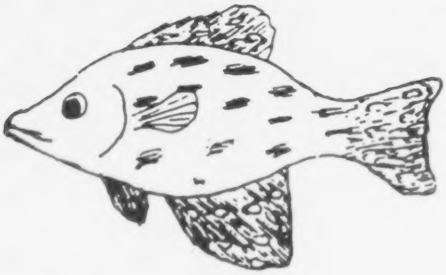
C.



Now explain why.

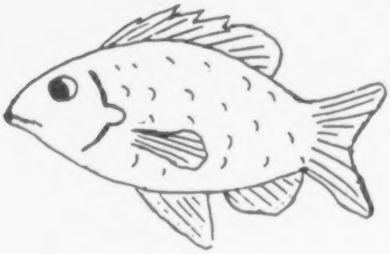
Letter B is differene
from all the outhers and
B is a catfish with
wishers and all the outhers
are differene from catfis.
The catfish whichs more
than the outhers. And the
catfish is biger

Sorting Fish
Sample 11

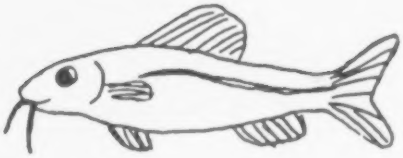


Which of the fish below is not like the above fish?

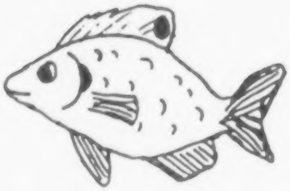
A.



[✓ B.]



C.



Now explain why.

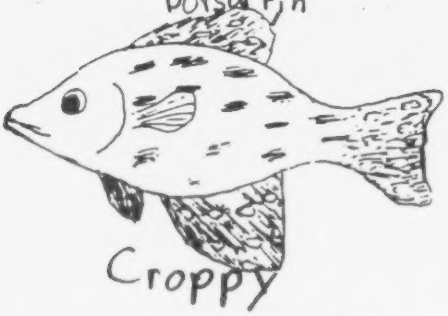
I picked fish B

It is different in a lots of ways.
This fish is a catfisn it has whiskers. The others don't. Fish B is not tall but it is fat. The others are tall but not fat. Fish B has teeth & is very tender. The others are rery hard & do not have teeth. Fish B hos spins. Fish A & C do not have spins.

A3.2,H2

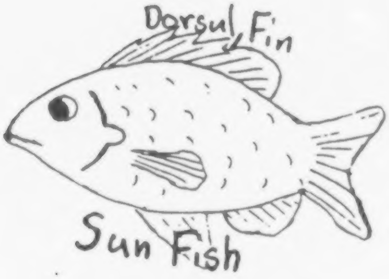
Sorting Fish

Sample 12

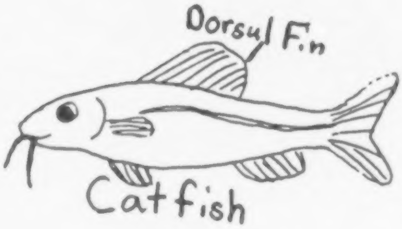


Which of the fish below is not like the above fish?

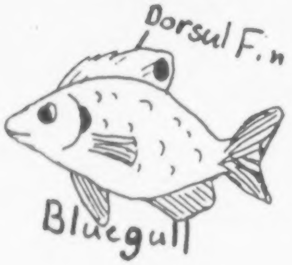
A.



[✓ B.]



C.



Now explain why.

The Catfish is different from the Croppy becuse it's dorsul fin is not jagid but the Croppy's is. The Catfish is more slender then the Croppy. The Catfish has a line down it's back not spots. The Catfich also has wiskers + the Croppy dose not.

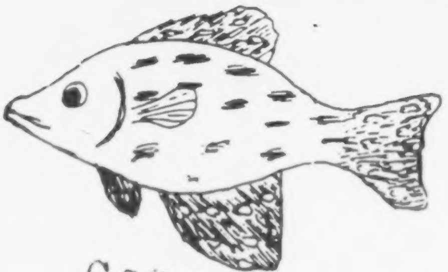
The Catfish is different from the Sun Fish because the Catfish has a line down it's back + the Sun Fish does not. The Catfish has no hump in it's gull iy cover. The Sun Fish has a jagid dorsul fin + the dorsul fin on the Catfish is not.

The Catfish is different from the Bluegull because the Bluegull has a jagid dorsul fin with a dot on it the Catfish dose not. The Catfish has a line down it's back the Bluegull doesn't. The Bluegull has a dot on it's gull cover the Catfish doesn't.

A3.2,H2

Sorting Fish

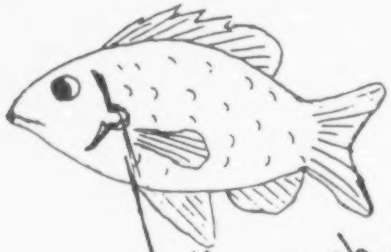
Sample 13



Crappie

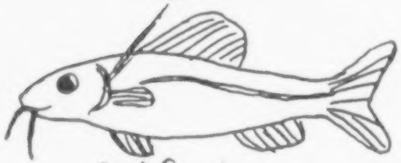
Which of the fish below is not like the above fish?

A.



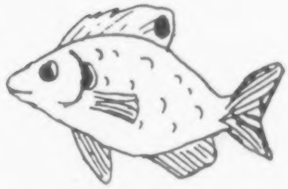
(is nush differental)

(B.)



Catfish

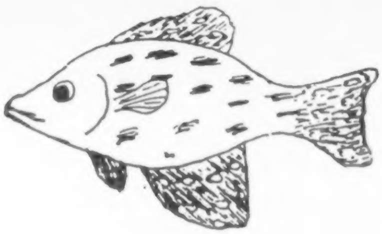
C.



Now explain why.

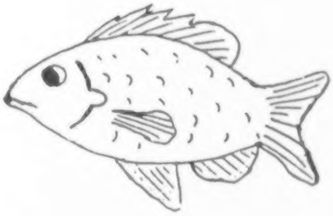
I think the reason why is because B has a wavey line down its backand no spots. It has no spines on the dorsal fin. It also has whisker mone like the other mush thines. the shape on B fish is alot different in shape.

Sorting Fish

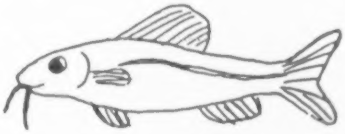


Which of the fish below is not like the above fish?

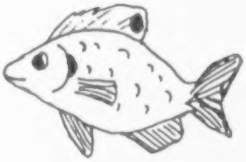
A.



B.



C.



Now explain why.

A3.2,O1

LABORATORY
NETWORK PROGRAM

Sorting Fish

Scoring Criteria (Task Specific)

- Ability to identify fish by name
- Use of descriptive language specific to the identification of fish
- Ability to make valid comparisons in differentiating between sample fish.
- Ability to follow directions, as indicated by comparison of sample fish to selected fish.

Scoring Rubric

3 points	Student names two or three of the kinds of fish. Student uses descriptive language to make multiple valid comparisons between the selected fish and the sample fish.
2 points	Student names at least one kind of fish. Student uses descriptive language to make at least one valid comparison between the sample fish and any other fish.
1 point	Student uses no descriptive language, but correctly identifies two or three of the different kinds of fish. OR, student uses descriptive language to make at least one valid comparison between any two or more fish, excluding the selected fish.
0 points	Student uses no descriptive language, does not identify any of the fish, and makes invalid or incoherent comparisons.

A3.2,O2

LABORATORY
NETWORK PROGRAM

Sorting Fish

Focus Questions

- What represents quality student work?
- What are the differences between developing or modifying criteria and rubrics *after* the task is administered to students, rather than *before*?
- What are the implications of engaging students in developing generalized performance criteria and rubrics, as you have just done?

A3.2,O3

LABORATORY
NETWORK PROGRAM

Activity 3.3

Miss Tolivar's Math Class

Purposes:

1. To illustrate the integration of assessment and instruction
2. To provide examples of alternative assessment in mathematics

Uses:

1. In any chapter to illustrate alternative assessment in mathematics
2. In Chapter 3 to illustrate the integration of assessment and instruction both as a tool for learning and a tool for continuous monitoring
3. In Chapter 4 to illustrate characteristics of good alternative assessment

Rationale:

To make sense of alternative assessment, and the relationships between curriculum, assessment, and instruction, it can be useful to have a visual model—particularly a model of these processes as they take place in an actual classroom. The *Good Morning Miss Tolivar* video provides an excellent example to initiate discussion of many of these issues. It is also useful in helping participants visualize assessment in a "new" mathematics classroom. The video *Good Morning Miss Tolivar* depicts scenes from the junior high mathematics classroom taught by Kay Tolivar of East Harlem Tech. Classroom scenes are interspersed with vignettes and comments from students and their relatives. Ms. Tolivar, a Presidential Awardee in mathematics, uses hands-on activities and communication skills in teaching mathematics.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- VCR and monitor

- The video, *Good Morning Miss Tolivar*, 1993, Foundation for Advancement in Science and Education[*]

- **Overheads A3.3,O1**—*Discussion of these activities*, **A3.3,O2**—*What messages?* and **A3.3,O3**—*What skills or concepts?*

- Chapter content to use as a handout: Continuous Monitoring

Time Required:

20-45 minutes

Facilitator's Notes:

The entire video is 27 minutes in length. If time is short, you may wish to show just the following segments:

- Opening classroom scene through "raisins" activity

- Opening classroom scene through "green slime" activity

- Opening classroom scene through "clothing" activity

You will want to include the short "Math Trail" segment that occurs near the end of the tape with all scenes.

1. After viewing selected sections of the video, ask participants to spend a few minutes writing their impressions of Ms. Tolivar's mathematics instruction.

2. Then ask participants to discuss one of the activities viewed with respect to the issues on **Overhead A3.3,O1**.

3. If you are particularly focusing on the consequences of assessment and/or bias in assessment, you can also discuss the questions on **Overhead A3.3,O2**.

4. If appropriate, you might also wish to explore the topic on **Overhead A3.3,O3**.

---

[*] The video may be obtained from: FASE Productions, P. O. Box 847, Los Angeles, CA 90078. Call: (800) 888-0600. Cost: $14.95 + $5.00 shipping and handling. (The video was sponsored by the National Science Foundation, ARCO, the U. S. Department of Energy, and the Corporation for Public Broadcasting. It may also be available from these agencies.)

Discuss one of the activities:

What are the purposes of Ms. Tolivar's questions during the activity?

What facets of the activity were assessment?

What facets of the activity were instruction?

In Ms. Tolivar's classroom, is it possible to separate assessment from instruction?

What types of evidence is Ms. Tolivar collecting about her students' learning?

What use is she making of this information?

A3.3.O1

LABORATORY
NETWORK PROGRAM

What messages about the nature of mathematics are being sent by Miss Tolivar's assessment/instruction?

- About how mathematics is done?
- About the contexts where mathematical knowledge is appropriate?
- About who can do mathematics?
- About who the students are?
- About the students' mathematical efforts?
- About what "counts" in mathematics learning?
- About the nature of problems and solutions in mathematics?

A3.3,O2

LABORATORY
NETWORK PROGRAM

What skills or concepts are being assessed/taught in the particular activities?

- raisins
- slime
- circles
- clothing
- bridges

A3.3,O3

LABORATORY
NETWORK PROGRAM

Chapter 4

Exploring Design Options for Alternative Assessment

What's in this Chapter?

While Chapters 2 and 3 strive to develop overall visions of the role of assessment in instruction and the implications of purpose for how assessments are designed, Chapter 4 begins the detailed examination of design options using specific science and mathematics alternative assessments as examples. What do some of the current assessments look like? When should we use various design options? How does assessment purpose relate to design options? These are the questions addressed in this chapter.

Chapter Goals

1. Illustrate design options for alternative assessments using material from current projects
2. Expand expertise in both evaluating and developing alternative assessments
3. Develop concepts used throughout the remaining chapters

The information in this chapter is relevant to both large-scale and classroom assessment.

Chapter Content

Readings

Background Information..Page 5

Our discussion of design options is based on having systematically collected and reviewed almost 200 alternative assessments in science and mathematics. This section describes how and why we gathered the assessments, so that readers can judge for themselves the basis for our conclusions. This section also develops an idea that is necessary for the remainder of the *Toolkit*—the notion of the difference between, and the importance of, both performance tasks and performance criteria.

Design Options—Tasks .... **Page 7**

Design Options—Performance Criteria .... **Page 13**

Other Observations on Current Assessments .... **Page 20**

Conclusion .... **Page 21**

References .... **Page 23**

Activities in this Chapter

Activity 4.1 *Performance Tasks—Keys to Success*

Illustrates the dimensions along which performance tasks differ. Discusses the advantages and disadvantages of different approaches. Time: 60 to 90 minutes

Activity 4.2 *Spectrum of Assessment Tasks*

Participants review a variety of short, related assessment tasks and rank them according to amount and type of information each might elicit from students. Time: 60 minutes

Activity 4.3 *Performance Criteria—Keys to Success*

Illustrates the characteristics of sound performance criteria. Time: 60 minutes

Related Activities in Other Chapters

Activity 2.5 *Clapping Hands*

Participants play the part of assessors and assessees to explore the consequences of designing alternative assessments in various ways. Time: 75 minutes

Activity 2.6 *A Comparison of Multiple-Choice and Alternative Assessment*

Participants compare a multiple-choice test to an alternative assessment that attempts to measure the same skills, and discuss the advantages and disadvantages of each approach. Time: 60 minutes

Activity 2.7 *Think of a Time*

Participants reflect on their own experiences being tested and draw conclusions about what they want their assessments to be like. Time: 20 minutes

Activity 2.9 *Going to School*

Demonstrates the importance of performance criteria, illustrates different kinds of performance criteria, and discusses the relative advantages and disadvantages of different types. Time: 30 to 90 minutes

Activity 2.10 *Tasks vs. Performance Criteria*

This discussion assists participants to look at some examples of alternative assessment and illustrates the two main parts of an alternative assessment—tasks and performance criteria. Time: 20 minutes

Activity 3.1 *Camping Trip: A Math Problem-Solving Task*

Demonstrates how to develop performance criteria and illustrates how the development of performance criteria can be instructional for both teachers and students. Time: 60 to 90 minutes

Activity 3.2 *Sorting Fish: A Science Task*

This activity is similar to Activity 3.1, but uses a science task instead of a math task. Time: 60 minutes

Activity 3.3 *Miss Tolivar's Math Class*

This activity uses a video that shows a teacher implementing continuous monitoring of student achievement. Time: 60 minutes

Activity 6.2 *How Can We Know They're Learning?*

A version of **Activities 3.1** and **3.2** designed for parents. Time: 2 hours

Sample Assessments Referred to in this Chapter

Sample assessments are used to illustrate the points made in this chapter. All samples, plus an index of descriptive information, are found in **Appendix A**. The following lists the samples that are referred to in this chapter.

Sample 4.1	*How Many Buttons?*
Sample 4.2	*Mathematics Problem Solving*
Sample 4.3	*Discovering the Problem of Solid Waste*
Sample 4.4	*Sow Bugs*
Sample 4.5	*Weathercaster's Helper*
Sample 4.6	*Aquarium Problem*
Sample 4.7	*Performance Assessment in Math (Alberta)*
Sample 4.8	*Science—New Directions in Assessment*
Sample 4.9	*Mapping the Blue Part*
Sample 4.10	*Writing in Chemistry*
Sample 5.1	*Assessment and Technology Videotape*
Sample 5.2	*Assessment of Laboratory Skills in High School Science*
Sample 5.3	*Science Portfolio (GSE)*
Sample 5.4	*Grade 3 Interdisciplinary Task*
Sample 5.5	*Assessment of Learning and Communication Processes*
Sample 5.6	*Algebra II*
Sample 5.7	*Mathematics Assessment*
Sample 5.9	*Primary Math Portfolio*

Exploring Alternative Assessment Design Options—Readings

Background Information[**]

What Types of Assessments Form the Basis of Our Analysis?

In order to illustrate a wide variety of approaches to alternative assessment design, the following broad definition of alternative assessment was adopted. This definition, first stated in Chapter 1, is repeated here for convenience.

> *Alternative assessment includes any type of assessment in which students create a response to a question rather than choose a response from a given list (e.g., multiple-choice, true/false, or matching). Alternative assessments can include short answer questions, essays, performances, oral presentations, demonstrations, exhibitions, and portfolios.*

Tasks vs. Performance Criteria

The major selection criteria for including an alternative assessment in the LNP-AA collection were that the instrument, technique, or procedure has two components:

[**] Assessments that formed the basis for the LNP-AA analysis came from many sources. Existing collections of assessment instruments at CRESST and NWREL were merged. Each regional educational laboratory was responsible for searching out other assessment initiatives in its own region. Labs followed a variety of search mechanisms from blanket mailings to targeted phone calls. At a minimum, every state department of education was contacted. Over two years, labs contacted over 1,300 individuals by mail or by phone.

Finally, there was additional collection at the national and international levels, including:

1. Examining journals from 1989 to the present
2. Searching databases such as ERIC, dissertation abstracts, Educational Index, Buros, and Tests
3. Sending letters and called over 100 colleges, universities, research centers, professional organizations, and other national and international sites
4. Sending letters to and did catalog searches of 16 publishers of educational assessments

After sample assessments were collected, they were screened using the criteria listed above and entered in the database. Complete entries were printed from the database and sent to authors for comment. Changes were made as needed. This process resulted in the assessments that formed the resource for the *Toolkit* and *Database*.

1. Specification of *tasks*. The *task* is the activity or assignment given the students to do. Student responses to the task are what is assessed. *Tasks* can be specific exercises or problems (for example, math problems, a specific laboratory experiment, a group project, or a portfolio), or specification of the circumstances under which information would be collected about students (for example, student cooperation might be observed during the course of regular lab activities in science).

2. *Performance criteria* or another specified way of evaluating student performance on the tasks given to them. Methods could include: right/wrong scoring, assignment of points for "correctness" of response, checklists of response features, and rating scales using professional judgment.

It is not assessment if these two components are not present.

Sometimes educators are confused about these components of an alternative assessment. To many, assessment means only the "task" (or problem, question, item, prompt, or activity) you give students to do. This is probably because we have thought about it this way for so long. On multiple-choice tests, for example, since scoring is right/wrong, the problem or question given the students *is* the major consideration. If the student gets the question correct then it is assumed that he or she knows whatever content or skill the question covered.

However, with alternative assessments the manner of scoring—the performance criteria—is as important as the design of the tasks when analyzing what the test assesses. For example, if the performance criteria only score conceptual understanding, it doesn't matter that the task requires problem solving, collaboration, making a presentation, or anything else—what is assessed is conceptual understanding.

Consider **Sample 4.7**—*Performance Assessment in Math (Alberta)*, "Bucket of Beans." In this task students have to estimate the number of beans in the bucket using various manipulatives. What this task actually assesses depends more on the criteria used to judge success than on the task itself. If it were scored right/wrong (i.e., the student gets it "right" if his or her estimation is within a certain value of the real number of beans) then one might assume that the test measures ability to estimate. However, the *actual* criteria for success focus on problem solving and communication. A high score is given not for getting the right answer, but for using good problem-solving strategies and communicating adequately what was done. Therefore, what is assessed is problem solving and communication, not estimation. The task of estimation using manipulatives is just the context for assessing problem solving and communication.

A common error when people are beginning to develop or choose alternative assessments is to only attend to development of the tasks. So you will hear, for example, people saying, "I do alternative assessment, I have students write journals," or "I do alternative assessment, I have students do computer simulations." But, the question remains, *What will you look for in those journal entries or in the response to the simulation that will tell you whether you are accomplishing your goals for students? What will you assess? How will you know when performance is good enough?* These questions are answered by the performance criteria.

Activity 2.10—*Tasks vs. Performance Criteria* provides examples of tasks and performance criteria so that individuals can distinguish these two parts of an alternative assessment. **Activity 2.9**—*Going to School* illustrates the importance of performance criteria when making consistent judgments about students. **Activity 3.1**—*Camping Trip*, **Activity 3.2**—*Sorting Fish*, and **Activity 6.2**—*How Can We Know They're Learning?* show the importance of performance criteria as instructional tools for students.

Design Options—Tasks

Performance *tasks* are the exercises or assignments that we give students to do; it is student performance on the *task* that is assessed. There are differences in how tasks are designed. These differences have implications for how science and mathematics education is conceptualized and for our philosophy of instruction. Therefore, it is useful to discuss some of the variations, so that you can relate the possibilities to your own instructional philosophy and goals for students.

In science and mathematics alternative assessments, the *Toolkit* authors found task design differences such as number of right answers, use of manipulatives, response mode, length/complexity, amount of group work, whether it is timed or untimed, and amount of student choice.

Open-Response vs. Open-Ended Tasks

In this *Toolkit* the term **open-response** refers to problems or tasks for which the student generates an answer, but there is still only one right, or a few best, answers. Thus, even though tasks may be complicated and require equipment or manipulatives, and even though we might assess student performance on the problem-solving process rather than getting the right answer, there is still only one right, or a few best, answers. All the examples in Figure 4.1 are open-response questions. (Relevant additional portions of assessment instruments and source information are provided in **Appendix A**—*Alternative Assessment Sampler*.)

Figure 4.1
Examples of Open-Response Tasks

A. 145 + 28 + 3767 + 90405 = __________ (school district test)

B. Estimate the number of beans in the bucket. (Students are given a large bucket of beans, magic markers, a tray and a small cup. **Sample 4.7**—*Performance Assessment in Math [Alberta]*)

C. Look at the weather symbols at the bottom of the page. Choose the weather symbol that best describes the weather today and draw a circle around it. (**Sample 4.5**—*Weathercaster's Helper*)

D. Examine the items in the garbage can. Sort the items into the following categories: paper, metal, food waste, plastic, yard waste, glass, other. Count and record the number of items in each category. List the categories in order from least number of items to greatest number of items. (The purpose of the exercise is to see whether students can classify different types of solid waste. This is open-response because *all* students are given a garbage can with the *same* trash in it and there is only one "right" way to sort it. **Sample 4.3**—*Discovering the Problem of Solid Waste*)

E. Do sow bugs prefer it where it is light or where it is dark? (Students are given experimental equipment. **Sample 4.4**—*Sow Bugs*)

F. Your task is to determine the location of specific features of the simulated ocean floor...Produce a graph which represents a profile of the ocean floor. (Students are given a black box with a simulated ocean floor and scale, and a height finder with a sliding ring. **Sample 4.9**—*Mapping the Blue Part*)

In the *Toolkit* the term **open-ended** refers to problems or tasks for which there is no single right answer; students could generate any number of acceptable responses. Consider the tasks in **Figure 4.2**.

Figure 4.2
Examples of Open-Ended Tasks

G. Your class will be getting a 30-gallon aquarium. The class will have $25.00 to spend on fish. You will plan which fish to buy. Use the *Choosing Fish for Your Aquarium* brochure to help you choose the fish. The brochure tells you things you must know about the size of the fish, how much they cost and their special needs. Choose as many different kinds of fish as you can. Then write a letter to me explaining which fish you choose. In your letter tell me how many of each kind of fish to buy, give the reasons you chose those fish, and show that you are not overspending and that the fish will not be too crowded in the aquarium. (**Sample 4.6**—*Aquarium Problem*)

H. Students produce a mechanized vehicle that produces at least two simultaneous motions in different directions to accomplish a set of actions. For example, a dump truck or an oscillating fan. (**Sample 5.1**—*Assessment and Technology Videotape*)

I. If you were an oceanographer wanting information concerning the ocean floor, would you use this method? Why or why not? What additional investigations would you use to map the ocean floor? (This is the second part to example F given above. **Sample 4.9**—*Mapping the Blue Part*)

J. What could you do to reduce the amount of solid waste in this garbage can? (This is the final part to example D shown above. **Sample 4.3**—*Discovering the Problem of Solid Waste*)

K. What do you think this might be the graph of? Put names and numbers on the graph to show what you mean. (**Sample 4.2**—*Mathematics Problem Solving*)

L. Write an essay in which you explain the most important ideas and principles that your friend should understand. (**Sample 4.10**—*Writing in Chemistry*)

There are many examples of both types of tasks in the materials obtained for the LNP-AA collection. One approach is not necessarily better than another. Open-response tasks seem to be associated with the attempt to measure knowledge of specific pieces of information or the ability to apply specific concepts to new situations, while open-ended tasks seem to be associated with the attempt to measure problem solving, group process, critical thinking, or science process skills.

This should not be taken to mean that this is what **you** should do. There are also assessments in which students are given open-response tasks, and student work is scored for problem solving (for example, **Sample 4.7**—*Performance Assessment in Math* [*Alberta*]). What you *should* do is have a clear idea of what you want to measure and a well-articulated instructional philosophy and approach, and then choose or design tasks that will fit your instructional philosophy and elicit the desired behavior on the part of students.

How Students Perform Tasks

There seem to be two major ways that students perform tasks—with and without the use of manipulatives or equipment. In the *Toolkit* the use of manipulatives or equipment is called "hands-on." Hands-on tasks contrast with assessments that are done only with paper and pencil—students manipulate symbols, draw pictures, or do calculations, all on paper. There are many examples of both types of procedures in the assessments collected. Examples of tasks from **Figures 4.1** and **4.2** that use manipulatives or equipment are: **B** (*Alberta Math*), **D** (*Solid Waste*), **E** (*Sow Bugs*), **F** (*Mapping the Blue Part*), and **H** (*Assessment and Technology Videotape*).

Examples from **Figures 4.1** and **4.2** that present tasks in paper and pencil format are: **A** (*Calculations*), **C** (*Weathercaster's Helper*), **G** (*Aquarium Problem*), **K** (*Mathematics Problem Solving*), and **L** (*Writing in Chemistry*).

Notice that both approaches (with and without the use of equipment and manipulatives) have been used to assess all kinds of outcomes: knowledge, skills, conceptual understanding, and "big" outcomes such as problem solving, communication, group process skills, and critical thinking. Again, one procedure is not necessarily better than another. Each technique has advantages and disadvantages. Users need to balance the following factors:

- purpose for the assessment
- the skills measured
- developmental level of students
- instructional philosophy
- what the assessment communicates to teachers and students about what is important
- practical necessity

Users need to be especially careful not to blindly use hands-on tasks because they seem to be a good idea. There has to be a good pedagogical reason to use a hands-on task. Likewise, we need to be aware of the messages we send to teachers, students, and parents by never using hands-on tasks.

Response Mode

The collected assessments also differ in how students respond. Most student responses are written (e.g., Samples **A** [*Calculations*], **B** [*Bucket of Beans*], and **C** [*Weathercaster's Helper*]; see also Samples **D**, **F**, **G**, **I**, **J**, **K**, and **L**); some are oral (e.g., **H** [*Assessment and Technology Videotape]*) and some require a physical response or production of a product (e.g., **E** [*Sow Bugs*] and **H** [*Assessment and Technology Videotape*]). (Note that responses could be multi-modal, such as example **H,** in which students produce a product and present an oral report about it.)

Written responses are more practical and produce a permanent record. Producing products also is fairly permanent, but bulkier, and might be a better way to assess some outcomes—for example, if you want to really see if a student can *do* something and not just talk about doing it. Oral responses might be more appropriate for younger students, students with poor writing skills, and students for whom English is a second language. *We need to be careful that ability to write does not mask the actual student skill being assessed.*

Length/Complexity

We're beginning to hear about the "three P's"—on-demand Performance assessment, Projects, and Portfolios. On-demand performance assessments are relatively short tasks (one to three class periods) that students do at more or less the same time in more or less the same manner. (That is, they tend to be more standardized.) Examples are: **B** (*Bucket of Beans*), **E** (*Sow Bugs*), and **F** (*Mapping the Blue Part*); see also **G, K,** and **L.** Another type of on-demand assessment is "constructed response" in which students provide short answers, usually to a series of related questions. Some authors distinguish this type of question from on-demand performance assessments, which tend to have longer, more integrated tasks. Example **C** (*Weathercaster's Helper)* illustrates constructed response questions.

Projects are longer term activities that can take the place of other instructional activities, are embedded in classroom work, and can occur more at the discretion of the teacher. An example is **H** (*Assessment and Technology Videotape*). Projects used for assessment are frequently indistinguishable from projects used for instruction. In fact, some people maintain that projects used for assessment should model good instruction. A good discussion of this can be found in Gong et al. (1992).

Portfolios are purposeful collections of student work that can show progress over time, depth of competency, etc. Samples in portfolios are frequently chosen by students to demonstrate what they know and can do, usually in pre-established areas. For example, **Sample 5.3**—*Science Portfolio (GSE)* is a high school science portfolio in which students select one problem-solving investigation, one piece of work that demonstrates creative expression, and one piece of work that demonstrates growth through writing for their portfolio, and explain why they were chosen. There is frequently much student choice in portfolio content. Thus, portfolios tend to look more individual. Standardization, when it occurs, usually occurs at the level of categories of things students must submit for their portfolios and in the criteria used to judge success, rather than at the level of mandating *particular* work or performances for inclusion.

Which of these options you select depends on what you want to assess and your purposes for assessment. For example, if you need to make high-stakes, important decisions about individual students, you need to gather sufficient evidence to be sure that your estimation of student ability is valid. Since some of the skills on which we judge competence might require complex behaviors, a portfolio might be called for. A portfolio allows the collection of multiple examples of complex performances over time. No single on-demand performance assessment allows this. A perfect example is the current Certificate of Mastery reform effort in several states in which students have to demonstrate mastery of complex abilities in order to receive an endorsement on their diploma. Attempts are being made to design portfolios for this purpose.

The point here is that task type is not just chosen because it is faddish or sounds like a good idea. Tasks must be crafted to accomplish what is desired. If you need a quick, overall estimate of how a group of students is doing on math problem solving, for example, an on-demand performance assessment of five problems randomly distributed among students (so that any single student only gets one problem) might be the best choice. If you need to make high-stakes decisions about individual students, a combination of approaches might be necessary.

Group Work

Some tasks are done individually (e.g., **E** [*Sow Bugs*], **G** [*Aquarium Problem*], and **L** [*Writing in Chemistry*]; see also samples **A**, **B**, **C**, **D**, and **K**), some are done in groups (**F** [*Mapping the Blue Part*] and **H**), and some are a combination (e.g., **Sample 5.4**—*Grade 3 Interdisciplinary Task)*. Group work is usually incorporated into tasks if:

1. You want to assess group collaboration. You probably don't want to design a separate task just to assess group collaboration. It is best done in the context of other tasks.

2. You want to send a message that group collaboration is important.

3. Students need to engage in group work to "level the playing field." For example, if you wish to assess ability to write, you might assign a task for which some students have more prior knowledge than others. Group work allows for

sharing background information that all students can then use to craft their essays.

4. Students need to interact in order to gather information needed for an individual task. For example, students might individually analyze survey data that is collected by the group.

Conclusion—Tasks

Tasks must be crafted to accomplish a vision. This vision relates to what you want to assess, the reason you are assessing it, the instructional messages to teachers you want to send, the messages to students you want to send, the relationship you see between assessment and instruction, and your instructional philosophy. This vision building is expanded on in Chapters 3 and 5.

Activity 4.1—*Performance Tasks—Keys to Success* and **Activity 4.2**—*Spectrum of Assessment Tasks* give participants in professional development sessions the opportunity to look at different task types and discuss their advantages and disadvantages based on the considerations outlined above. **Activity 2.5**—*Clapping Hands* is a compelling way to have participants generate conclusions about how tasks should be designed. **Activity 2.6**—*A Comparison of Multiple-Choice and Alternative Assessment* asks participants to compare two types of tasks. **Activity 3.3**—*Miss Tolivar's Math Class* demonstrates tasks that are embedded in instruction.

Design Options—Performance Criteria

In alternative assessments the quality of student performance on a task is assessed or rated using some sort of *performance criteria, rubric* or *scoring guide*. In the science and mathematics alternative assessments collected by the LNP-AA, the following seem to characterize the types of scoring mechanisms found.

Task-Specific vs. Generalized Criteria

Probably the major distinction between scoring types relates to the degree that scoring is task-specific. **Task-specific** scoring means that the scoring guide can only be used for one, single task because you look for certain specific features in order to award points. One form of task-specific scoring is marking answers right or wrong as in **Sample 4.3**—*Discovering the Problem of Solid Waste*.

Scoring for **D** (**Sample 4.3**—*Solid Waste*)

Sorting: 7 points for all correct. Subtract 1 point for each incorrect response. Sequencing: 1 point for ordering items correctly; 0 points for incorrect order.

Or, scoring can be based on features of responses, as in **Sample 4.1**—*How Many Buttons?*:

Scoring for **Sample 4.1**—*How Many Buttons?*

Students work together to estimate the number of buttons on people's clothing in the school. A preliminary activity had students estimate the number of buttons in their classroom. A *high response* takes into consideration the following aspects of the problem: (a) the data collected from the day before; (b) some estimate of the number of children per classroom; (c) the number of classes in the school; (d) variation of buttons among children of various ages; and (e) adults' buttons. Reasonable justifications are given for how each numerical value was chosen. Numerical values are put together with appropriate arithmetic processes. Steps are clearly detailed.

Neither of these scoring guides could be used for another task. For example, you couldn't use the scoring guide for **Sample 4.1** to score the student responses in **Sample 4.3**. Task-specific scoring is also used in **Samples 4.5**—*Weathercaster's Helper* and **4.6**—*Aquarium Problem*.

Now consider the following **generalized** criteria:

Scoring for **K** (**Sample 4.2**—*Mathematics Problem Solving*)

Performance is scored along four dimensions: conceptual understanding, procedural knowledge, problem solving/strategies, and communication. The same scoring guide is used for all problems given to students in grades 3, 5, 8, and 11.

Scoring for **H** (**Sample 5.1**—*Assessment and Technology Videotape*)

Performance is scored for thinking skills (conceptual understanding, creativity, and critical thinking), group collaboration, and communication skills (speaking and writing). These same criteria could be used for any group project.

Other generalized criteria appear in **Sample 4.7**—*Performance Assessment in Math (Alberta)*, **Sample 4.8**—*Science—New Directions in Assessment*, and **Sample 4.10**—*Writing in Chemistry.*

This difference in the design of performance criteria relates to the degree to which one infers competence *indirectly* from performance on a task versus judging it *directly*. Although the importance of this may not be entirely clear at first, it really does have far-reaching implications for how one views and uses alternative assessment. As an illustration, let's assume that we are designing an alternative assessment to measure mathematics problem solving.

The **indirect inference** approach would consist of developing a math problem for students to solve that the developers believe requires problem solving. Then points are assigned for different features of the students' performance: getting a particular answer, having a particular graph with particular labels, using particular descriptive phrases, having a particular picture, using particular measuring instruments in particular ways; in short, task-specific scoring. Problem-solving ability is not judged directly, but inferred from the number of points the student receives. Thus, task-specific scoring is often associated with indirect approaches of assessing skills. For example, a high score on **Sample 4.6**—*Aquarium Problem* is taken to be evidence of problem-solving ability.

Other good examples of this are **Sample 4.5**—*Weathercaster's Helper,* in which students are assigned points for the correctness of various answers and then these points are added up in various ways to obtain subtest scores for four science process skills, and **Sample 4.3**—*Discovering the Problem of Solid Waste,* where high scores are taken as evidence that students understand the concepts related to solid waste.

This is essentially the same logic as in multiple-choice tests, the only differences being that (a) with alternative assessments the answer is generated by the student and not chosen from a list, and (b) multiple-choice tests usually are scored right/wrong, while answers to alternative assessment tasks can be given any number of points depending on particular features in addition to the right answer.

Contrast this to the other end of the continuum for our problem-solving illustration—**direct judgment** of the skill being assessed. In this approach we would design tasks to elicit problem solving, but then, instead of scoring the response right/wrong or assigning points for correctness or specific features, we would rate problem solving directly using **generalized criteria**. For example, consider the four-trait model developed by the Oregon Department of Education (**Sample 4.2**). In this case, the *same* criteria are applied to *all* problems solved by students and problem solving is judged directly by the rater.

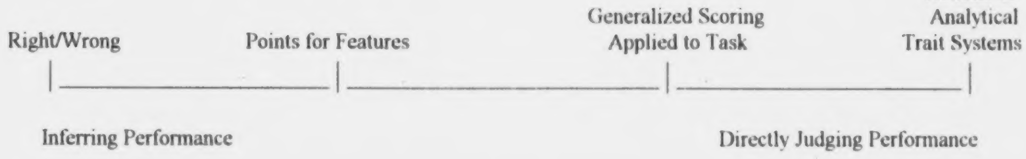
This difference is important because it relates to our vision of how alternative assessment fits into the educational enterprise. For example, the indirect approach (emphasizing task-specific scoring) is more reliable and faster for large-scale assessment because raters don't have to analyze each student's work for evidence of problem solving; all they have to do is look for specific diagrams, calculation or explanations, and give points for their presence. All the judgment work has already been done—someone else has already decided that on Problem A, good problem solvers would have this diagram, calculations, or explanation.

A more direct approach (emphasizing more generalized scoring procedures), however, has important advantages for instructional use in the classroom because it *forces assessors to do the analysis themselves*. Making judgments about the qualities that make a performance successful helps assessors internalize the criteria for success. This becomes especially powerful if the assessors are teachers and students. (See **Activity 3.1**—*Camping Trip*, **Activity 3.2**—*Sorting Fish* and **Activity 6.2**—*How Can We Know They're Learning?* for demonstrations of the power of this approach.) The following sentiments (although addressing writing and not problem solving) explain the link to instruction.

> *Scales, criteria and specific questions which students apply to their own or others' writing also have a powerful effect on enhancing quality. Through using the criteria systematically, students appear to internalize them and bring them to bear in generating new materials even when they do not have the criteria in front of them. [This procedure is] two times more effective than [other] techniques.*
> (Hillocks, 1986)
>
> *Winning points may be the final goal of classroom work as it is of the sports endeavor, but the grade, like the final score of the game, never taught anyone how to win again, or why they lost. For the truly successful contenders, playing the game is always about learning the game...however often it seems to be about scoring more wins than losses.* (Lucas, 1992)

In other words, making direct judgments of the skill being assessed helps teachers and students conceptualize "what good looks like" across situations. When scoring is done for specific features of tasks, raters only get to know what "good" looks like on a single problem; they don't have to generalize, transfer or apply knowledge to a new situation. Therefore, direct rating of performance using generalized criteria is better as an instructional tool.

Of course, life is never black and white. What has been described are two ends of a continuum:


Right/Wrong
Points for Features
Generalized Scoring Applied to Task
Holistic or Analytical Trait Systems
Inferring Performance
Directly Judging Performance

In actuality, alternative assessment approaches fall all along this continuum. One "happy medium" can occur in cases where a general set of scoring guidelines is developed and then systematically applied to each task to develop specific scoring guides based on points. This approach can actually help teachers and students generalize so that eventually they aren't dependent on specific scoring guides. Such an approach is used by Kentucky (**Sample 4.9**—*Mapping the Blue Part*).

This may be a good compromise when large-scale efficiency is required or when users (including students) are learning how to apply general criteria to varied tasks. To be successful, however, the general criteria have to adhere to standards of quality (comprehensive and descriptive); the link between the specific example (e.g., "has a picture showing a whole cut into four equal parts") and the general guide (e.g., "uses appropriate graphic representations") has to be clear; and the ultimate goal has to be understanding the general criteria and applying them to new situations.

Holistic vs. Analytical Trait Generalized Scoring

Within the broader category of generalized criteria, there are also two approaches—holistic and analytical trait. **Holistic** scoring entails giving one overall judgment about the quality of the response as a whole. **Analytical trait** scoring entails giving separate scores for various dimensions of performance. **Sample 4.8**—*Science—New Directions in Assessment* is a good example of generalized holistic scoring of science process skills. To get the highest score the student must demonstrate a clear understanding of the problem, make valid observations, design a well thought-out experiment, present data well, support conclusions with facts, etc. Other examples of holistic scoring are **Sample 4.9**—*Mapping the Blue Part*, **Sample 5.7**—*Mathematics Assessment*, and **Sample 5.9**—*Primary Math Portfolio*.

Sample 4.2—*Mathematics Problem Solving* is a good example of generalized analytical trait scoring for mathematical problem solving. Each student response receives four judgmental scores—understanding the problem, procedural knowledge, problem solving/strategies, and communication. Other examples of analytical trait scoring are **Sample 4.7**—*Performance Assessment in Math (Alberta)*, **Sample 4.10**—*Writing in Chemistry*, **Sample 5.1**—*Assessment and Technology Videotape*, **Sample 5.2**—*Assessment of Laboratory Skills in High School Science*, **Sample 5.5**—*Assessment of Learning and Communication Processes*, and **Sample 5.6**—*Algebra II*.

The advantage of holistic scoring is that it is quicker. Therefore, it is used in large-scale assessment when the main goal is to get a relatively quick overview of student performance. The advantage of analytical trait scoring is that it is more diagnostic. Therefore, it is used when the goal is either to provide diagnostic information to teachers to plan instruction or when the goal is to use criteria as a way for students to understand and internalize the criteria for success.

Nature of the Scoring Scale

Once you have decided on the major approach to scoring (task-specific or generalized) and decided whether you will have one score or several, you need to decide how those points will be awarded. To make this clear, consider the holistic generalized criteria used in **Sample**

4.8—*Science—New Directions in Assessment.* First assessors had to decide they would assign a maximum of four points (and not 2, 3, 5, 6, 7, 8, or whatever). Then they had to decide what would constitute a score of 4, 3, 2, or 1. In other words, what features make a performance excellent, good, medium or poor?

Believe it or not, there *are* variations in the philosophy behind how points are assigned for each trait scored. Here are some:

Scoring Based on Quantity

In **Sample 4.10**—*Writing in Chemistry,* students are scored on four traits—concepts, argumentation, use of references to the text, and misconceptions. Three of these are based on "counts." That is, you count up the number of misconceptions, concepts used, or proportion of text detail used, and the score depends on the number you find.

Scoring Based on Quality

In **Sample 4.8**—*Science—New Directions in Assessment* judgments are based on the quality of the responses. For example, one feature of a high performance is that "the student successfully proposes an explanation which clearly shows a relationship between data and final conclusions." The corresponding feature of a weak performance is that "the final conclusion is unrelated to the tasks performed." It doesn't matter how many conclusions the student comes to as long as the conclusions drawn are good ones.

In general, rating scales based on quantity should be avoided because, for example, one good conclusion should be worth more than 10 poor conclusions.

Amount of Descriptive Detail

Scoring with Sketchy Descriptions of Terms

In **Sample 4.3**—*Discovering the Problem of Solid Waste,* to get a high score for "Solution" (see the bottom of the final page of the sample), the student must have given a "strong, workable solution." There is no assistance with what it means to have a strong, workable solution. What does a strong, workable solution look like?

Scoring with Some Descriptions of Terms

In **Sample 4.2**—*Mathematics Problem Solving,* to get a high score on "Procedural Knowledge," the student "uses principles efficiently while justifying the solution, uses appropriate mathematical terms and strategies, solves and verifies the problem, and uses mathematical principles and language precisely." While still not completely definitive, this provides additional assistance with what to look for in the response in order to help make a judgment.

Scoring with More Complete Descriptions of Terms

In **Sample 5.3**—*Science Portfolio (GSE)*, consider the additional detail provided in order to help the assessor rate the student performances he or she will encounter. For example, in the "Growth Through Writing" rubric, a strong score is indicated when "Student work and self-reflection show extensive knowledge of the scientific ideas presented. The student uses relevant terms appropriately, providing evidence of excellent growth in understanding the scientific ideas through comparison of original and revised samples of work..." and so on.

There are two issues related to amount of detail. First, it is easier for raters to be consistent when description increases. This is especially important for large-scale assessment. However, an even more important consideration than just getting good consistency between assessors is the communication potential of descriptive criteria. If we tell students, for example, that their solutions have to be "workable" will that help them know what to do? It would be more helpful if we were to say that "workable" means a balance between being practical (in terms of time, knowledge, and money) and being effective (in terms of really solving the problem); that effectiveness has to be demonstrated by more than just opinion (you need data on the effectiveness of other similar solutions); and that the technology needs to be currently available to implement the solution. Can you think of additional details that would be useful for students?

The point here is that good quality performance criteria can be more than just a tool for giving a score on a test. They can help us define our expectations for students and develop a shared vision of our goals for students. (See **Activity 3.1**—*Camping Trip*, **Activity 3.2**—*Sorting Fish,* and **Activity 6.2**—*How Can We Know They're Learning?* for exercises that illustrate the descriptive and communicative potential.)

Conclusions—Performance Criteria

There are many types of performance criteria under development. The kind of scoring mechanism directly relates to what you want to assess, your purpose for assessment, and how you see alternative assessment fitting into the educational enterprise. Holistic and task-specific scoring are often used for large-scale assessment because they are faster. Analytical trait performance criteria having lots of description may be the best choice when you want to develop a common vision of the goals for students or if you want to maximally increase the chances that alternative assessments will be useful instructional tools in the classroom. The themes of (1) creating a vision of assessment (a tool for monitoring student progress or a tool for instruction) and (2) purposes for assessment (especially large-scale vs. instruction) will be revisited throughout this *Toolkit* as we examine assessment design considerations.

As a side note, it is interesting that, while almost all large-scale assessments have attended to developing reliable scoring mechanisms, many teacher-developed assessments tend to be sketchy in this area. Many assessments were rejected from inclusion in the LNP-AA collection because they had no performance criteria at all!

Activities in the *Toolkit* related to performance criteria include:

- **Activity 2.9**—*Going to School* lets workshop participants try out different kinds of performance criteria and discuss their advantages and disadvantages. It also demonstrates the necessity of having performance criteria for large-scale assessment.

- **Activity 4.3**—*Performance Criteria—Keys to Success* talks about the essential features of performance criteria if they are to be used as tools for communication in the classroom.

- **Activity 3.1**—*Camping Trip*, **Activity 3.2**—*Sorting Fish*, and **Activity 6.2**—*How Can We Know They're Learning?* illustrate how to develop performance criteria and show how they can be tools for defining our goals for students.

Other Observations on Current Assessments

Level of Skills Addressed

Assessments in the LNP-AA collection vary widely on what they try to measure: facts, concepts, application of facts or concepts, science process skills, or "big" outcomes such as problem solving, communication skills, critical thinking, group process, and reasoning.

Some likely examples of each type are shown below. These are "likely" because what any task really requires on the part of the student depends on what the student was taught; a recall task for a high school physics student might be a problem-solving challenge for a sixth-grader. The point with the examples below is *not* that there is one *correct* interpretation of content, but rather that users think about what they are trying to assess in order to pick the task types that might be best.

Examples of assessing facts from **Figures 4.1** and **4.2** are examples **A** (*Calculations*) and **C** (*Weathercaster's Helper*). Examples of attempts to assess conceptual understanding and application of concepts are examples **D** (*Solid Waste*—applying knowledge of types of garbage to an actual task), **F** (*Mapping the Blue Part*), and **H** (*Assessment and Technology Videotape*—applying knowledge of mechanics). Examples of assessing science process skills are **E** (*Sow Bugs*) and **F** (*Mapping the Blue Part*—using efficient processes). Attempts to measure general process skills are examples **B** (*Bucket of Beans*—communication and problem solving), **L** (*Writing in Chemistry*—persuasive writing), and **G** (*Aquarium Problem*—problem solving). (Also, please note that our judgment of what is assessed is based on examining both the task and the performance criteria.)

The emphases in the alternative assessments examined by the LNP-AA is measuring conceptual knowledge, application of concepts, and general process skills because alternative assessments can be expensive and time-consuming. "Alternative assessment dollars" need to be spent wisely—on what is most difficult to assess in another format and what essential outcomes would be missed if an alternative assessment was not used.

Technical Information

Technical information on instruments covers such things as whether the tasks elicit the correct student performance, how consistently responses can be scored, whether tasks or scoring are biased toward any particular group, whether the results can be generalized to other similar tasks, and the impact of the assessment on students, teachers, curriculum, and instruction (see Chapter 5 for additional discussion of these topics). In general, except for well-funded large-scale assessment projects, very little technical information is available on the instruments collected by the LNP-AA. In some cases, pilot testing is under way. In other cases, little or no attention seems to have been paid to this important topic. When technical information is available, it seems to be mostly descriptive—only average scores of groups.

Conclusion

There is considerable activity in the development of alternative assessment in science and mathematics at all levels of education (classrooms to federal governments) and at all grade levels. A variety of approaches is evident in the samples the LNP-AA has collected and analyzed. In all these examples, the same reasons are given for changes in assessment: the need to better reflect current theories of instruction, changes in instructional goals for students, etc. However, the link between the general rhetoric and the specificities of individual assessment designs is not often clear. For example, if everyone is saying we need to better model good instructional practice, why do some assessments only use manipulatives while others only use traditional word problems? Or, why, when people are assessing math problem solving, do some groups use task-specific scoring while others are developing general criteria? We need to step back and reconsider all our design features to make sure they really do match what we are trying to accomplish. The remainder of this *Toolkit* will also assist in this process.

References

Clarke, D. (1993). A continuum of task types. Victoria, Australia: Mathematics Teaching and Learning Centre, Australian Catholic University.

Gong, B., Venezky, R. & Mioduser D. (1992). Instructional assessments: Lever for systemic change in science education classrooms. *Journal of Science Education and Technology, 1*, pp. 157-176.

Hillocks, G. (1986). *Research on written compositions.* Urbana, IL: National Council of Teachers of English.

Lucas, C. (1992). Introduction: Writing portfolios—changes and challenges. In K. Yancey (Ed). *Portfolios in the writing classroom*, pp. 1-11. Reston, VA: National Council of Teachers of Mathematics, Inc.

ACTIVITIES

Activity 4.1

Performance Tasks—Keys to Success

Purposes:

1. To illustrate the dimensions along which performance tasks differ
2. To discuss the characteristics of quality performance tasks
3. To discuss the advantages and disadvantages of different approaches in light of the purpose of the assessment, instructional philosophy, and vision of integrating assessment and instruction

Uses:

1. In Chapter 2 or 4 to illustrate different task types
2. In Chapter 5 to discuss the advantages and disadvantages of different types of tasks

Rationale:

A variety of performance task types are being developed—everything from short answers that look like multiple-choice with the choices taken off to complex, long term, group projects. These differ in complexity, the degree to which they require integration of skills or test individual skills in isolation, amount of collaboration, length of time required, presentation and response modes, amount of student choice, and the number of correct answers. It is necessary for educators to consider which of these options will accomplish their assessment purposes.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Handout **A4.1,H1**—*Criteria For Performance Tasks* and Overhead **A4.1,O2**—*How Tasks Differ*
- Sample performance tasks; one set of possible choices is: **Sample 4.2**—*Mathematics Problem Solving*, **Sample 4.5**—*Weathercaster's Helper*, **Sample 5.1**—*Assessment and Technology Videotape*, **Sample 5.4**—*Grade 3*

Interdisciplinary Task, **Sample 5.8**—*A Day at the Carnival*, and the multiple-choice test that is part of **Activity 2.6**—*A Comparison of Multiple-Choice and Alternative Assessment;* run these off as handouts as well as prepare them as overheads

Use the list below if you want to use samples to illustrate specific design features. For brevity, we have only listed the title of the sample the first place it is mentioned.

Multiple-choice: Use the test example in **Activity 2.6**—*A Comparison of Multiple Choice and Alternative Assessment*
Short Answer: Samples 4.5—*Weathercaster's Helper*, **5.8**—*A Day at the Carnival*, **4.3**—*Discovering the Problem of Solid Waste*, **5.4**—*Grade 3 Interdisciplinary Task.*
On-Demand Performance Assessment: Samples 4.1—*How Many Buttons?*, **4.2**—*Mathematics Problem Solving*, **4.3**, **4.4**—*Sow Bugs*, **4.6**—*Aquarium Problem*, **4.7**—*Performance Assessment in Math [Alberta]*, **4.8**—*Science—New Directions in Assessment*, **4.9**—*Mapping the Blue Part*, **4.10**—*Writing in Chemistry*, **5.2**—*Assessment of Laboratory Skills in High School Science*, **5.4**, **5.6**—*Algebra II*, **5.7**—*Mathematics Assessment*
Project: Samples 5.1—*Assessment and Technology Videotape*, **5.5**—*Assessment of Learning and Communication Processes*
Portfolios: Samples 5.3—*Science Portfolio(GSE)*, **5.9**—*Primary Math Portfolio*

One right answer: Samples 4.3, 4.4, 4.5, 4.7, 4.9, 5.6, 5.8
More than one right answer: Samples 4.1, 4.2, 4.6, 4.9, 4.10, 5.1, 5.3, 5.4, 5.5, 5.9

All paper and pencil: Samples 4.2, 4.5, 4.6, 4.10, 5.7, 5.8
Use of manipulatives: Samples 4.1, 4.3, 4.4, 4.7, 4.8, 5.1, 5.2, 5.4

All individual work: Samples 4.2, 4.4, 4.5, 4.6, 4.7, 4.8, 4.10, 5.2, 5.7, 5.8
Some group work: Samples 4.1, 4.9, 5.1, 5.4, 5.5, 5.6

Testing skills in isolation: Samples 4.1, 4.3, 4.5, 5.4, 5.8
Assessing skills used together: Samples: 4.2, 4.7, 4.8, 4.9, 4.10, 5.1, 5.2, 5.3, 5.5 5.6, 5.9

Time Required:

60-90 minutes

Facilitator's Notes:

1. As needed, review what performance tasks are. (Also see **Activity 2.10**—*Tasks vs. Performance Criteria.)*

   We use the term to refer to whatever activity project, prompt, problem, or assignment is given to the students to do. The quality of the student response to the task is what is assessed. As needed, brainstorm examples of tasks.

   The main purpose of the task in an alternative assessment is to elicit the right performance while avoiding pitfalls—things that can cause us to make an incorrect judgment of student skill level. By "eliciting the right performance" we mean that students do what you want them to do. Here are some examples of times that the task did not elicit the right performance. (Come up with your own examples, or use the examples below.)

   a. On a third grade writing assessment, the students were given the prompt: "You've heard the story about how the camel got its hump. Think of something in nature and write a story about how it got that way." The intent was to assess imaginative writing. What was actually obtained were thousands of student essays on why camels have humps. Imaginative writing could not be assessed because the students wrote informational pieces instead.

   b. Would the camping trip problem in **Activity 4.2** assess problem solving if students had just completed several problems just like it? In this case, the problem might be eliciting recall of information, not problem solving.

   The point here is that if you want to assess critical thinking, for example, you have to have a task that results in students thinking critically. Or, if you want to assess, for example, group process skills, you have to give students a task that enables them to work cooperatively.

   Ask the participants to provide examples from their own classrooms of tasks or activities that did not work as intended.

2. *For many of the outcomes we want to assess with alternative assessment, we need "meaty" tasks or the right performance will not be elicited.* To get this point across, ask participants to think of a successful activity they have used with students and what made it so. Common responses at all grade levels are: choice, active, clear instructions, engaging, relevant, organized, students know the purpose, enough time, several ways to approach the task, and accommodates different learning styles.

   The interesting thing about the list that teachers generate is that this is the same list used by many authors to illustrate what is meant by "authentic" assessment. For example, see Handout **A4.1,H1**—*Criteria for Performance Tasks*. Point this out to participants. Good assessment tasks that will elicit the right performances from students look a lot like good instructional tasks.

3. Use **A4.1,H1**, the "meatiness" scale, to evaluate and discuss several performance tasks. Use examples you have discussed previously, or, if this is a stand-alone activity, a thought-provoking set of tasks are:

   a. The fake multiple-choice test in **Activity 2.6**—*A Comparison of Multiple-Choice and Alternative Assessment*. Participants usually rate this type of test low on the "meatiness" scale.

   b. **Sample 4.5**—*Weathercaster's Helper*, or **Sample 5.8**—*A Day at the Carnival.* These are usually rated fairly low on the "meatiness" scale.

   c. **Sample 5.4**—*Grade 3 Interdisciplinary Task.* Although short answer with lots of right/wrong responses, this task is usually rated in the middle of the "meatiness" scale because it is more hands-on.

   d. **Sample 4.2**—*Mathematics Problem Solving*. This graphing problem is all paper and pencil, but is still fairly open-ended and engaging. Participants usually rate it about two-thirds to three-quarters up the "meatiness" scale.

   e. **Sample 5.1**—*Assessment and Technology Videotape*. This group project is usually rated very high on the "meatiness" scale.

   These task types can be though of as "anchors" for the "meatiness" scale. (Remember good criteria have samples to illustrate the range of quality. If **A4.1,H1** represents criteria for tasks, then it is important to have samples of tasks that illustrate the range of task quality.)

4. Go back over the task examples to see what features tend to make them more or less "meaty." Handout **A4.1,H2**—*How Tasks Differ* can be used to focus the discussion.

5. The first characteristic of a good task is to elicit the right performance. This has been the focus of steps 1-4 above. The second characteristic of a good task is to avoid pitfalls—sources of mismeasurement. Some of these probably came up during the discussion of tasks above. Participants may have said things like:

   a. "This requires a lot of writing. What about students who don't write well?"

   b. "This requires a lot of reading. What about students who are ESL?"

   c. "Will students who have had experience with carnivals do better than students who haven't?"

   d. "How will group work affect the judgment of ability of individuals?"

When these comments arise, note that this is what is meant by "sources of mismeasurement" and keep a running list. Then, at this point, have participants brainstorm additions to the list. You could even categorize the list into major sources of mismeasurement: equity, fairness, and bias; extraneous interference; sampling.

6. An extension would be to have participants attempt to develop a performance task that would elicit a valued student skill.

CRITERIA FOR PERFORMANCE TASKS

Here is a list, adapted from one developed by Steve Leinwand and Grant Wiggins, of some important characteristics that we should look for in tasks. We will find it helpful to return often to this list, reviewing the quality of the curriculum and assessment activities we present to students.

ESSENTIAL	• The task fits into the core of the curriculum. • It represents a "big idea."	vs.	TANGENTIAL
AUTHENTIC	• The task uses processes appropriate to the discipline. • Students value the outcome of the task.	vs.	CONTRIVED
RICH	• The task leads to other problems. • It raises other questions. • It has many possibilities.	vs.	SUPERFICIAL
ENGAGING	• The task is thought-provoking. • It fosters persistence.	vs.	UNINTERESTING
ACTIVE	• The student is the worker and decision maker. • Students interact with other students. • Students are constructing meaning and deepening understanding	vs.	PASSIVE
FEASIBLE	• The task can be done within school and homework time. • It is developmentally appropriate for students. • It is safe.	vs.	INFEASIBLE
EQUITABLE	• The task develops thinking in a variety of styles. • It contributes to positive attitudes.	vs.	INEQUITABLE
OPEN	• The task has more than one right answer. • It has multiple avenues of approach, making it accessible to all students.	vs.	CLOSED

"Criteria for Performance Tasks," from *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions*, edited by Jean Kerr Stenmark. Reston, VA: National Council of Teachers of Mathematics, 1991. Reprinted with permission.

A4.1,H1

How Tasks Differ

- One right answer vs. more than one
- Written, oral, and/or visual instructions
- Written, oral, and/or visual activities
- Written, oral, and/or visual responses
- Length: On-demand, project, portfolio
- Complexity: Atomistic or integrated
- Individual or group work
- Timed or untimed
- Student choice
- What else?

A4.1,H2

Activity 4.2

Spectrum of Assessment Tasks

Purposes:

1. To understand that assessment tasks provide "snapshots" of students' knowledge and understandings from which we make inferences about their abilities
2. To understand that altering assessment tasks in small ways can open them up and move them along the assessment spectrum to provide truer pictures of what students know and can do

Uses:

1. In Chapter 2 to review the advantages and disadvantages of various task formats
2. In Chapter 4 to design tasks that will elicit desired outcomes
3. In Chapter 5 to evaluate tasks for equity, accessibility and quality

Rationale:

In this activity, participants review a variety of short, related assessment tasks and rank them according to the amount and type of information each might elicit from students. Opportunities to review a range of task types help participants make distinctions between subtle characteristics of tasks that affect their usefulness and appropriateness in assessments. By noting slight changes in context as well as content and the way in which questions are asked, participants learn ways that tasks can be modified to be more powerful.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Chart paper and marking pens
- Sheets of chart paper displaying the questions from the opening activity

- **Overheads A4.2,O1**—*Mathematics Multiple Choice Item,* **A4.2,O2**—*Science Multiple Choice Item,* **A4.2,O3**—*Ask Yourself...,* and **A4.2,O4**—*Assessment Spectrum*

- **Handouts A4.2,H1**—*Mathematics Exercises/Tasks* and **A4.2,H2**—*Science Exercises/Tasks*

Time Required:

60 minutes

Facilitator's Notes:

A. Overview (5 minutes)

1. Briefly introduce yourself and, if the group is not too large, have participants do likewise. You may wish to do the Opening activity (see B below) and then the rest of the Overview (A).

2. Provide the following brief overview of the session goals:

   In this activity, you will review a variety of short, related assessment tasks and rank them as to the amount and type of information each might elicit from students. The goals are to understand that:

   - *Assessment tasks provide "snapshots" of students' knowledge and understandings, from which we make inferences about their abilities.*
   - *Altering assessment tasks in small ways can open them up and move them along the assessment spectrum to provide truer pictures of what students know and can do.*

B. Opening (5-10 minutes)

1. Show either the multiple choice item for math (**A4.2,O1**—*Mathematics: Multiple-Choice Item)* or for science (**A4.2,O2**—*Science: Multiple-Choice Item)*. If you are doing both math and science, you will repeat this part of the activity using the other overhead. (Note: These items appeared on a recent test from the National Assessment of Educational Progress. They were obtained from the ASCD Newsletter, *Update*, March 1993, p. 5.)

2. Ask the following questions, posting each at the top of a sheet of chart paper, and record participants' responses on the chart paper. *What would a student need to know in order to answer this item? What would you know about the*

student who responded correctly? What would you know about the student who responded incorrectly?

3. Tell participants that the item recently appeared in the National Assessment of Educational Progress exam administered to 12th graders nationwide. Provide the NAEP statistics as to the percent of 12th graders who responded correctly—math item: 45 percent correctly answered **A.** Science item: 62 percent correctly answered **D.**

4. Ask the following question and allow participants a few minutes to discuss and share their ideas as a large group: *What would you do if these results were from your class?*

C. Background (10-15 minutes)

1. Share the following background information with participants:

   Assessment tasks are opportunities for students to show what they know and can do. They help "paint the picture" of the many dimensions of a student's understandings and misunderstandings. Although someone might use a hammer to drive a screw or use a screwdriver to hammer a nail, we usually try to match our tools to the purpose of our tasks. Assessments are tools we use to gauge what students know and can do. It is important to choose the correct assessment tools so that we can make appropriate inferences from a student's performance.

2. Display the overhead *Ask Yourself* (**A4.2,O3**), and continue with the following information:

   To determine what questions or tasks to ask of your students, ask yourself, "What do I want to know about my students' knowledge and understanding?" and "What will I do with the results?" As you consider and create possible tasks, ask "Will the task(s) elicit useful responses so that I ***can*** *assess my students' knowledge and understanding?" Just as instructional tasks provide students with opportunities to learn, tasks used for assessment must provide teachers with the opportunity to assess.*

   Consider the NAEP multiple-choice item(s). Certainly, tasks of this sort may provide information about what students might know, but what insights will a teacher have of what the same students can do?

   If you are using this activity in connection with Chapter 2 (Supporting Educational Improvement with Alternative Assessment), emphasize that we are examining characteristics of good assessment, clarifying our ideas of what to assess while keeping in mind a clear purpose for assessing. This activity helps provide the rationale for assessment alternatives. In Chapter 5 (Evaluating the

Quality and Equity of Alternative Assessments), the activity helps clarify what makes a good task.

3. Display only the top section of the overhead *Assessment Spectrum* (**A4.2,O4**), keeping the bulleted lists covered with a sheet of paper for the time being. Tell participants:

   We will be using an "Assessment Spectrum" to gauge where a variety of related assessment tasks might fall, indicating a range of involvement that tasks might elicit from students. Concepts tested with traditional, single–solution tasks can often be "moved along" the Assessment Spectrum towards the more alternative assessments by "opening up" the questions asked. For example, we can reword questions so that more than one solution exists, or so that there are a variety of ways to reach a solution. And, we can have students show one or more of those alternatives.

4. Move the sheet of paper on the overhead to reveal the bulleted items, allowing time for participants to read through some ways in which traditional tasks differ from alternatives.

5. Refer back to the NAEP item and ask participants: *Where along the Assessment Spectrum—from "traditional" to "alternative"—might the NAEP item(s) be located?* Accept ideas from a few, but allow no more than a brief discussion at this time.

6. Tell participants:

   Depending upon the students' level of expertise and familiarity with particular types of problems, alternative assessments typically have increased cognitive complexity. They challenge students to consider additional possibilities or aspects of a situation while analyzing given information, making and testing conjectures, and reconsidering implications of a chosen solution or path.

   The sample tasks in the following Assessment Spectrum activity show how small changes in the questions asked can provide bigger pictures of what students know and can do.

7. Distribute the handout of sample tasks in math, **A4.2,H1**—*Mathematics Exercises/Tasks*, and/or science, **A4.2,H2—***Science Exercises/Tasks.* Tell participants:

   On your own, read through the tasks and decide where you would place each one on the spectrum from "traditional" to "alternative." For our purposes, consider "alternative" tasks to be ones in which (1) the student is more actively engaged with the target knowledge and understandings to be assessed, and (2) opportunities exist for students to show what they know and can do in a variety of ways.

Allow a few minutes for individuals to read through the tasks and place each task on the spectrum.

D. Spectrum Decisions (20 minutes)

Have participants work in small groups of three to five with others having the same sample tasks. Give each group chart paper or blank transparencies and markers. Tell participants:

Without discussion, each of you share your Assessment Spectrum with the others in your group. Tell what characteristics of the task you considered in making your decisions.

Then, work with your group to reach consensus at least on the order, if not the exact locations, the tasks might be placed on the spectrum. Consider the questions asked in the opening activity. (See chart paper.) Take notes on your areas of agreement and disagreement, and what task characteristics were most important in deciding. Prepare to present your ideas to the large group.

E. Presentations (10 minutes)

1. Have each group present its Assessment Spectrum. Have participants briefly summarize similarities and differences among their final task locations, as well as areas of agreement and disagreement about task characteristics.

2. Ask and discuss:

   What ideas from this activity might you use in planning assessments and using the results with your own students so you get a true picture of your students' knowledge and understanding?

3. End the session by reminding participants of the following:

   Remember, with any single task, you create only a "snapshot" of a student's knowledge and understandings. To get accurate pictures of your students, you want the best snapshots possible and you need a large number and variety of them.

   The best of cameras with the best of lenses can take poor pictures if used incorrectly. But a poor camera with a poor lens can rarely take a good picture. Even if used with the best of intentions, poorly constructed or inappropriate tasks can create false pictures of students' knowledge and understanding. Choose your tools wisely and use them appropriately.

F. Extensions

1. To expand the range of the assessment spectrum, have participants consider additional task types, including Fermi problems, projects, and investigations. Those interested in similar mathematics tasks at the high school or college levels may also find ideas in *A Continuum of Task Types* by David Clarke (from the Mathematics Teaching and Learning Centre, Australian Catholic University, Victoria).

2. At some point, participants may wish to discuss how to go about using more open-ended tasks such as the ones in this activity, those in the handout (see above), or those in other chapters of this *Toolkit*. They may also want to consider why these forms of assessment tasks might be troublesome for students. David Clarke suggests the following:

   Steps for Using Open-Ended Tasks

   a. Generate lots of tasks

   b. Give to students

   c. Examine weeks later asking the question: What do I know about my students?

   Clarke also reminds us there are a variety of reasons why students may not do so well on such tasks:

   Difficulties With Open-Ended Tasks

   a. Great reluctance to give more than a simple answer

   b. Incorrect inference because of linguistic problems or absence of context

   c. Their conception of what constitutes an acceptable response

   A discussion with participants may provide more insights into the challenges and potential rewards facing those who wish to try alternative assessment tasks.

Throughout, have participants keep in mind these words about assessment (also found on **A6.2,O1**):

> ***Assessment is about the exchange of information.*** *Assessment is not primarily about the collection of information—but rather about its use in communicating about learning. Information flows both ways. We have to respond to what the student creates and how s/he responds. Assessment is NOT simply a tool for saying "Bad luck, you get it wrong."* —D. Clarke, Sept. 28, 1993

Mathematics Exercises/Tasks

Where would you locate each item along the Assessment Spectrum?

Traditional ↔ **Alternative**

A

The perimeter of a square is 36 cm. What is its area?

81 cm^2 36 cm^2

144 cm^2 1296 cm^2

B

What is the area of a square that has a perimeter of 36 cm? Show how you know.

C

A shape has a perimeter of 36 cm. What could its area be?

D

What is the perimeter of a square that has an area of 36 cm^2? Show how you know.

E

A rectangle has a perimeter of 36 cm. What could its area be? Draw pictures to show your thinking.

F

Draw or describe a shape that has a perimeter of 36 units and covers the greatest possible area. How would you convince a classmate it covers the greatest possible area?

G

Draw several different shapes that have a perimeter of 36 units. Which covers the greatest area? Tell how you know.

H

What are the length and width of a rectangle that has a perimeter of 36 cm and covers the greatest possible area? How did you decide?

I

Which is the better fit: A square peg in a round hole or a round peg in a square hole? How did you decide?

J

Make an object using 36 identical cubes. Draw it. Find its surface area. Compare your work with other students. What are the minimum and maximum possible surface areas for objects built with all 36 cubes? Explain your response.

A4.2,H1

Science Exercises/Tasks

Where would you locate each item along the Assessment Spectrum?

Traditional
Alternative

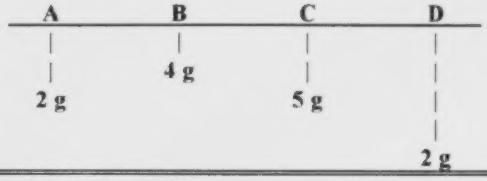
A

Suppose you measure the *periods* (the time it takes for a pendulum to swing back and forth) of two pendulums. Pendulum A has a mass of 5 g and pendulum B has a mass of 10 g. Each pendulum has a length of 50 cm. How will the periods compare?

- **A.** The period of pendulum A will be twice as long as the period of pendulum B.
- **B.** The period of pendulum B will be twice as long as the period of pendulum A.
- **C.** The periods will be the same.
- **D.** None of the above.

B

Suppose you are experimenting to find out if changing the weight on the end of a pendulum changes its period—the amount of time the pendulum takes to swing back and forth. Which of the pendulums shown below could you use for the experiment? Explain your reasoning.



C

You are given a set of four different pendulums. Two of the four pendulums have the same mass. Two are the same length, but do not have the same mass. Which of your four pendulums will swing back and forth at the same speed? Illustrate the four pendulums. Explain your answer and refer to your diagram.

D

Your design team has been asked by the City Park Department to construct a model for a new playground near the elementary school. The playground will have swing sets and seesaws. For the safety of the children who will be using the playground equipment, you must design your swings so that they don't swing too fast or loop–the–loop over the top of the swing set. Design and conduct an experiment to determine how the variables—length, mass, height of release—affect the rate of back–and–forth movement of a swing. Be prepared to present your findings, your recommendations, and a demonstration to the City Park officials.

E

Suppose you measure the periods of two pendulums. One pendulum has a mass twice that of the other pendulum. Also, one pendulum has a length twice that of the other pendulum. How might the periods compare? Use diagrams to help explain the possibilities.

F

Design and conduct an experiment to show how changing the weight on the end of a pendulum affects the period of the pendulum.

A4.2,H2

Mathematics
Multiple-Choice Item

The perimeter of a square is 24 centimeters. What is the area of that square?

a. 36 square cm

b. 48 square cm

c. 96 square cm

d. 576 square cm

e. I don't know

A4.2,O1

LABORATORY
NETWORK PROGRAM

Science Multiple-Choice Item

A scientist develops a theory to explain some phenomena that previous theories could not. However, this theory leads to predictions that are contrary to other scientists' expectations.

What should be done in response to these results?

a. Ignore the expectations and accept the theory.

b. Reject the theory since it is contrary to the expectations.

c. Revise the theory so that it agrees with the expectations.

d. Design experiments to test for the predictions made by the theory.

e. Develop another theory that predicts what the scientists expected.

A4.2,O2

LABORATORY
NETWORK PROGRAM

Ask Yourself...

What do I want to know about my students' knowledge and understanding?

What will I do with the results?

Will the task(s) elicit useful responses so that I *can* assess my students' knowledge and understanding?

A4.2,O3

LABORATORY
NETWORK PROGRAM

Assessment Spectrum

Traditional	Alternative
Fill-in-the-bubble	Open-ended tasks
Recognition, rote skills	Higher-order thinking or complex skills
Artificial and abstract situations	Mirror real-life situations
Paper and pencil	Performance-based
Efficient	Significant time and resources
Isolated student work	Individual or group performance
Programmed response	Student choice
Machine scoring, answer keys	Rater (human) scoring, scoring rubrics

Where would you locate the item along the Assessment Spectrum?

A4.2,O4

LABORATORY
NETWORK PROGRAM

Activity 4.3

Performance Criteria—Keys to Success

Purposes:

1. To illustrate the characteristics of sound performance criteria—what do good performance criteria look like?
2. To demonstrate the additional features of performance criteria that make them most useful as instructional tools in the classroom
3. To summarize previous points about the relationship of the assessment purpose to the design of performance criteria

Uses:

1. In Chapter 4 to discuss the advantages and disadvantages of different kinds of performance criteria
2. To follow-up **Activity 2.9**—*Going to School* and **Activity 4.1**—*Camping Trip*
3. An alternative to the performance criteria portion of **Activity 5.1**—*How to Be a Good Consumer*

Rationale:

A variety of performance criteria types are being developed. The type to use depends on: (1) the type of student outcome being assessed; (2) whether the assessment is large-scale or classroom; and (3) the vision one has of the role of assessment in the educational process. These issues are discussed in the text in Chapters 4 and 5. The purpose of this exercise is to demonstrate, with specific examples, the characteristics of performance criteria most useful when the focus is classroom assessment and the goals being assessed are the "big" outcomes of problem solving, critical thinking, communication skills, collaborative working, etc.

Materials:

For all options:

- Overhead projector, screen, blank transparencies, transparency pens
- Detailed analytical trait criteria for math problem solving, e.g. Oregon's 4-trait rubric, **Sample 4.2**—*Mathematics Problem Solving*
- Detailed holistic criteria, e.g., California's **Sample 5.7**—*Mathematics Assessment*
- **Overhead A4.3,O1**—*Performance Criteria—Keys to Success*

In addition, choose *one* of the following:

Option 1:

- Two "name the graph" student solutions, **A4.3,O2**—*Number of Teeth;* **A4.3,O3**—*Fruits* (These can also be used at handouts.)
- **A4.3,O5**—*Task Specific Criteria, Version 1: Name the Graph* and **A4.3,O6**—*Task Specific Criteria, Version 2: Name the Graph* (These can also be used as handouts.)

Option 2:

- Two student solutions from **Activity 2.9**—*Going to School* (**Handout A2.9,H2**)
- **A2.9,H3**—*Task-Specific Scoring Rubric* and **A2.9,H4**—*Generalized Scoring Rubric*

Option 3:

- Develop your own set of materials. You will need (1) two student responses—one strong and the other weak, (2) relative quality type criteria, and (3) detailed criteria.

Time Required:

60 minutes

Facilitator's Notes:

In many groups it is necessary to do other activities prior to this one in order to develop enough perspective to have a meaningful discussion. One good sequence for beginners is described in Chapter 7—Six to Seven Hour Training Event.

1. *The first "key to success"* **(A4.3,O1)** *for performance criteria is: Don't leave out anything of importance.* Describe why it is essential to cover everything of importance and give examples. You might say something like: *Well-defined performance criteria are a statement of what we value in student work. They communicate to students what to do and what is important to attend to. If we leave out important dimensions of performance, it is likely that they won't be emphasized because they won't be assessed. An example in writing is a state that left the trait of "voice" out of their rubric because they felt it would be too difficult to get good rater agreement. The result—no systematic attention to voice in writing. Another example is in music. What if you were a fiddle player and bowing wasn't rated as part of performance? What message does that communicate? Or, if teachers were held accountable for student performance, what effect might leaving bowing out have on instruction? A third example might be the Olympics. What if entry into the water was not taken into account in diving? Instead,* type of bathing suit *is scored. What message does this send about what is important? How would it affect coaching?*

   After sharing some examples from real life to illustrate the points made above about communicating what is valued, show the strong student performance (**A4.3,H1**—*Number of Teeth*). Ask, *How well did this student communicate what he or she did?* Most teachers agree that the student did a good job on this trait. Now put up the *Skimpy Criteria* (**A4.3,O4**) and point out the section on communication. Then cover up the trait of communication and ask, *What message would be communicated if communication was not scored? How might instruction be affected?*

2. The second key to success (**A4.3,O1**) is having criteria that are clear and detailed with samples of student work. Refer to the *Skimpy Criteria* (**A4.3,O4**). Have participants score the strong and/or weak performances on the trait of *communication*. Chances are that rater agreement will be pretty good. Now, pose the dilemma: *What if you give feedback using this scale and the student says, "I know I communicated well because my score is high, but I don't know why. Why did I get a good score on communication?"*

   The point to make is that students have a harder time generating another strong response if they don't know what made the previous response strong. The point of detail is not necessarily rater agreement (although detail will help here too); rather, detail helps to clearly communicate to students what they need to do to be successful.

3. The third key to success (**A4.3,O1**) is using generalized performance criteria. Define "generalized" and "task specific" and discuss examples. You could say: *Generalized means that the criteria can be used to rate a whole class or category of performances and not just the work from a specific task. Task-specific criteria can only be used to rate performance on a single task. For example, in the Olympics the same criteria are used to rate all dives—difficulty, form, entry into the water, etc.; there are not separate criteria for each dive. In writing assessment, criteria are often very general—some places use the same criteria to rate all kinds of writing; others have separate rubrics for persuasive, narrative, and expository writing. But, very few are task-specific, i.e., can only be used for a single topic.* Now discuss the relative advantages and disadvantages of these types of criteria. **Activity 2.9**—*Going to School* also includes such a discussion.

   The value of generalized criteria is that they help us to define what "good" looks like so that we can begin to generalize across performances and bring information from past experiences to bear on new experiences. This is especially important for the "hard to define" or "disagreement on what this means" goals such as critical thinking, problem solving, collaborative working, communication skills, etc. If people already agree on the definitions, then you can cut corners. But, people don't often agree on what these mean and what it looks like when students do "it" well. We can't hold students accountable for different visions of the same target. It is our moral obligation to define precisely what we mean.

   Thus, generalized criteria are best for classroom uses when "big" outcomes are being assessed. They are also, obviously, necessary when you will be looking at work generated by students on different tasks. This is the case for portfolios. They might not be required when one is assessing application of concepts as in **Sample 4.3**—*Discovering the Problem of Solid Waste. However, even generalized criteria for conceptual understanding are possible. Consider the following from* **Sample 5.1**—*Assessment and Technology Videotape*:

   > *The student understands relationships between variables and topics covered; locates pertinent information to solve problems; explains known principles, concepts, and theories and how they fit into work; uses examples to demonstrate knowledge; and uses relevant terminology.*

   Another good way to get these points across is to have participants look at task specific criteria and the problem and ask, *What does this assess?* You can't tell. With task-specific scoring, the procedure is frequently to classify questions by the skills they are supposed to assess and then add up total points for all questions that cover this skill. The total points is an indication of how "proficient" the student is on this skill. You never really have to define the skill in question at all. (See **Sample 4.5**—*Weathercaster's Helper* or **5.8**—*A Day at the Carnival* for good examples of this procedure. Also, **Sample 4.1**—*How Many Buttons?* shows that, when doing task-specific scoring, you need information in addition to the scoring guide to know what is being assessed.)

A third way to get this point across is to ask participants to score one of the problems in **Activity 3.1**—*Camping Trip* with the task-specific criteria on **A4.3,H4**—*Task Specific Criteria, Version 1: Name the Graph*. You can't do it. Then ask them to score the graph on **H4.3,H2**—*Fruits* with the generalized criteria in **Sample 4.2**—*Mathematics Problem Solving*. This *can* be done. This illustrates what "generalized" means.

Have participants score the problems with the various performance criteria types and discuss the advantages and disadvantages of each procedure. The advantages are that task-specific and relative quality criteria are quick, have good reliability, and are fast to learn. The disadvantages are you can't tell what's being assessed without additional backup material, a new rubric has to be developed for each task, and they doesn't generalize to the next task.

The advantages of generalized rubrics that are detailed are increases in understanding of what we mean by, for example, good problem solving; if teachers are scorers their expertise will improve, if students are scorers it is an instructional tool; different rubrics do not have to be developed for each problem; generalizable; can tell what is being assessed just from the criteria; a direct measure of a skill. The disadvantages are that it takes longer to learn.

(Note: Participants can more easily "see" the value of general performance criteria with writing. Draw parallels between writing and math problem solving. For example: *To learn writing performance criteria, we examine many samples of student writing in order to gain an understanding of how the traits look in different samples. The same is true of math. In math, you may at first want to explicitly convert the general performance criteria into task-specific performance criteria so people can see how to think about it. For example, one trait might be "conceptual understanding." How, on a specific problem, would conceptual understanding look?)*

4. The fourth key to success (**A4.3,O1**) is analytical trait vs. holistic criteria. Define those terms. For example, you might say, *Analytical trait means that more than one dimension of performance is assessed. Holistic means that you make one overall judgment of the quality of the response.* (**Caution**: Some people use these terms differently. "Analytic" for folks in California, for example, means "task specific" points, and "holistic" means *any* judgmental process. So, be sure participants understand that our use of both these terms implies judgment. The only difference is the number of judgments to be made—one [holistic] or several [analytical trait]).

   Have participants score the student performances with the analytical trait and holistic rubrics and discuss advantages and disadvantages. Holistic advantages are:

quicker, easier to learn, quicker rater agreement. Holistic disadvantages are: two vastly different papers can get the same score for different reasons. Analytical trait advantages are: provides better diagnostic information, provides better feedback to students. Analytical trait disadvantages are: more to learn, takes longer.

Performance Criteria—
Keys to Success

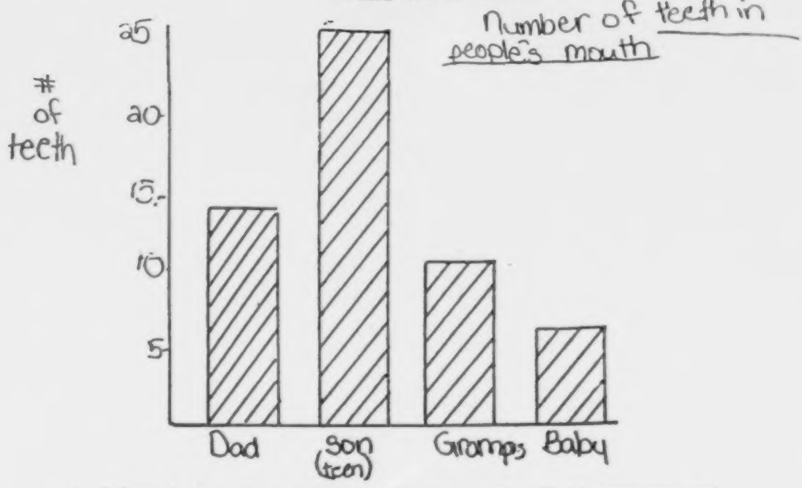
- **Don't leave out anything of importance** What they see is what you'll get
- **The more detail the better** Rich description, illustrated with samples of student work
- **Use generalized rather than task-specific criteria**
- **Consider analytical-trait** (several ratings) **rather than holistic** (one rating)

A4.3,O1

LABORATORY
NETWORK PROGRAM

Graphing Problem - Grade 5

NAME THE GRAPH



1. What do you think this might be the graph of? Put names and numbers on the graph to show what you mean.

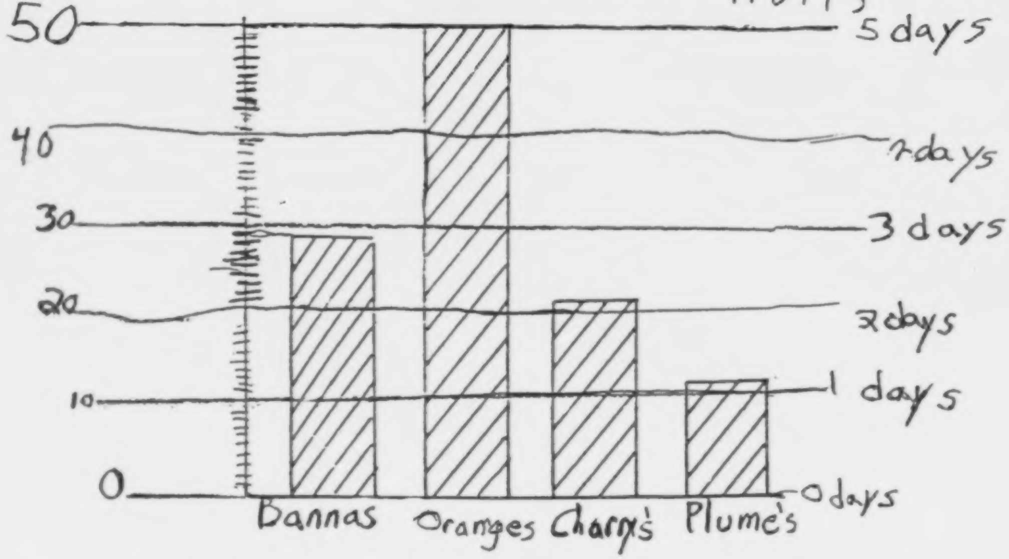
2. Write down everything you know from your graph.

1. Son has more teeth
2. Baby has the least
3. Gramps has more teeth than Baby.
4. Dad has more teeth than Gramps & Baby.
5. Son has about 25 teeth.
6. Dad has about 15 teeth.
7. Gramps has about 10 teeth.
8. Baby has about 6 teeth.
9. The name of the graph is # of teeth in people's mouths
10. The graph is about Baby, Gramps, Dad, and Son.
11. On this graph the most teeth you can have is 25.
12. Son has more teeth than Baby.

A4.3,H1
A4.3,O2

LABORATORY NETWORK PROGRAM

NAME THE GRAPH fruits



1. What do you think this might be the graph of? Put names and numbers on the graph to show what you mean.

2. Write down everything you know from your graph.

One day Mike ate 12 plumes. The next day Mike ate 21 charrys. The next day Mike ate 29 Banna's. The next day Mike ate 5 oranges. If you add them up he ate 112 In all.

A4.3,H2
A4.3,O3

LABORATORY
NETWORK PROGRAM

Skimpy Criteria

	Frequently	Occasionally	Sometimes	Never
Understanding the problem	4	3	2	1
Appropriate use of math	4	3	2	1
Clearly focused, good reasoning	4	3	2	1
Effective communication	4	3	2	1

A4.3,H3
A4.3,O4

LABORATORY
NETWORK PROGRAM

Task-Specific Criteria, Version 1: Name the Graph

4: Has a title under each bar, a label for each axes, and appropriate values listed on the y-axis. Has a title for the whole graph. The values along the y-axis are spaced appropriately and relate to the relative heights of the bars. The labels under each bar make sense. There are at least five things listed for the second question.

3: The student graph is essentially correct, but has at least one major thing wrong. For the most part, conclusions in part 2 are reasonable—at least three of the five statements are correct.

2: The student shows some understanding of the sorts of things that might go into a graph, such as the need for labels and numbers, but the response has serious errors, such as: labels that don't make sense; and numbers on the y-axis that don't bear a relationship to the height of the graphs. Conclusions are sketchy.

1: Most parts of the graph display lack of understanding, are missing, or are just plain wrong. Conclusions do not logically follow from the graph.

A4.3,H4
A4.3,O5

LABORATORY
NETWORK PROGRAM

Task Specific Criteria, Version 2: Name the Graph

Number of Points	**Feature**
1	Graph titled
1	x-axis labeled
1	y-axis labeled
1	Bars labeled
2	Values on y-axis spaced appropriately and related correctly to bars
3	1 point for up to two correct conclusions 2 points for three to four correct conclusions 3 points for five or more correct conclusions

A4.3,H5
A4.3,O6

LABORATORY
NETWORK PROGRAM

Chapter 5

Evaluating the Quality and Equity of Alternative Assessments

What's in this Chapter?

It is critical that educators become good consumers and developers of alternative assessment tools. Not only does assessment have a major impact on the quality of instruction, assessments are becoming increasingly visible and the stakes of assessment are growing as assessment becomes more widely used as a political and social tool. Thus, it is important that both classroom teachers and those designing large-scale assessments have guidelines with which to judge the assessment instruments they use.

Chapter 5 summarizes, and in some cases expands on, the quality issues and discussions introduced in previous chapters. Considered are content and skills coverage, the quality of performance tasks and criteria, equity and fairness, the consequences of assessment, and cost-effectiveness. Cross-references to previous chapters and activities help users address these somewhat technical topics in the way most suited to their own experience and background. Included are a rating form for alternative assessments and several equity activities.

Chapter Goals

1. Present guidelines for assessing the quality and appropriateness of an alternative assessment for particular purposes and contexts

2. Provide experience in choosing good assessments

3. Examine issues of equity, fairness, bias, and unintended consequences of assessment

Chapter Content

Readings

Criteria for Evaluating Alternative Assessments .... Page 6

This section defines reliability and validity, discusses the importance of considering these issues, and presents guidelines for evaluating alternative assessments. Guidelines include: (1) content and skill coverage, (2) performance criteria, (3) performance tasks, (4) fairness and bias, (5) consequences of assessment, and (6) cost and efficiency. Each topic is covered in detail.

Further Readings.. Page 19

Activities in this Chapter

Activity 5.1 *How to Be a Good Consumer of Science/Mathematics Alternative Assessments—or—How to Tell the Gems From the Duds*

This activity gives participants practice with making judgments about the quality of the science and mathematics alternative assessments they may want to try. Areas of examination include coverage, performance tasks, performance criteria, equity and fairness, consequences, and cost. Time: 6 hours

Activity 5.2 *Dr. Know-It-All*

This video-based activity illustrates with humor many subtle and not-so-subtle biases and messages sent by instruction and assessment activities in science. Discussion questions focus on how issues raised by the video may apply to the assessments participants are examining. Time: 30 minutes

Activity 5.3 *Chickens and Pigs: Language and Assessment*

This awareness activity emphasizes the critical role that language plays in effective and equitable assessments. Time: 15 minutes

Activity 5.4 *Equity Role Play*

Participants explore their personal beliefs, attitudes, and values by creating and playing the role of a student or teacher involved in potentially inequitable assessment activities. Time: 90 minutes

Activity 5.5 *Questions About Culture and Assessment*

This activity is designed to increase awareness of the relationship between culture and assessment and demonstrate how cultural factors can affect student ability to show what they know and can do. Time: 90 minutes

Other Resources in This Chapter

Alternative Assessment Evaluation Form (In **Activity 5.1**)

A rating form that can used to evaluate the quality of alternative assessments. The categories on the rating form parallel the issues discussed in the chapter.

Related Activities in Other Chapters

Activity 2.5 *Clapping Hands*

Participants play the part of assessors and assessees to explore the consequences of designing alternative assessments in various ways. Time: 75 minutes

Activity 2.6 *A Comparison of Multiple-Choice and Alternative Assessment*

Participants compare a multiple-choice test to an alternative assessment that attempts to measure the same skills, and discuss the advantages and disadvantages of each approach. Time: 60 minutes

Activity 2.7 *Think of a Time*

Participants reflect on their own experiences being tested and draw conclusions about what they want their assessments to be like. Time: 20 minutes

Activity 2.9 *Going to School*

Demonstrates the importance of performance criteria, illustrates different kinds of performance criteria, and discusses the relative advantages and disadvantages of different types. Time: 30 to 90 minutes

Activity 2.10 *Tasks vs. Performance Criteria*

This discussion assists participants in looking at some examples of alternative assessment and illustrates the two main parts of an alternative assessment—tasks and performance criteria. Time: 20 minutes

Activity 3.1 *Camping Trip: A Math Problem-Solving Task*

Demonstrates how to develop performance criteria and illustrates how the development of performance criteria can be instructional for both teachers and students. Time: 60 to 90 minutes

Activity 3.2 *Sorting Fish: A Science Task*

This activity is similar to Activity 3.1, but uses a science task instead of a mathematics task. Time: 60 minutes

Activity 4.1 *Performance Tasks—Keys to Success*

Illustrates the dimensions along which performance tasks differ. Discusses the advantages and disadvantages of different approaches.
Time: 60 to 90 minutes

Activity 4.2 *Spectrum of Assessment Tasks*

Participants review a variety of short, related assessment tasks and rank them as to the amount and type of information each might elicit from students.
Time: 60 minutes

Activity 4.3 *Performance Criteria—Keys to Success*

Illustrates the characteristics of sound performance criteria. Time: 60 minutes

Activity 6.2 *How Can We Know They're Learning?*

A version of **Activities 3.1** and **3.2** designed for parents.
Time: 2 hours

Sample Assessments Referred to in this Chapter

Sample assessments are used to illustrate the points made in this chapter. All samples, plus an index of descriptive information, are found in **Appendix A**. The following lists the samples that are referred to in this chapter.

Sample 4.1	*How Many Buttons?*
Sample 4.2	*Mathematics Problem Solving*
Sample 4.3	*Discovering the Problem of Solid Waste*
Sample 4.4	*Sow Bugs*
Sample 4.5	*Weathercaster's Helper*

Sample 4.6	*Aquarium Problem*
Sample 4.7	*Performance Assessment in Math (Alberta)*
Sample 4.8	*Science—New Directions in Assessment*
Sample 4.9	*Mapping the Blue Part*
Sample 4.10	*Writing in Chemistry*
Sample 5.1	*Assessment and Technology Videotape*
Sample 5.2	*Assessment of Laboratory Skills in High School Science*
Sample 5.3	*Science Portfolio (GSE)*
Sample 5.4	*Grade 3 Interdisciplinary Task*
Sample 5.5	*Assessment of Learning and Communication Processes*
Sample 5.6	*Algebra II*
Sample 5.7	*Mathematics Assessment*
Sample 5.9	*Primary Math Portfolio*

Evaluating the Quality and Equity of Alternative Assessments—Readings

Criteria for Evaluating Alternative Assessments

Introduction

Alternative assessment does not automatically ensure good assessment. Merely looking "interesting" or "authentic" is not sufficient to assure quality. It is perfectly possible to do a poor performance assessment, design a poorly conceived and executed portfolio system, or produce assessment activities that do not address valued outcomes. Two terms commonly used when discussing the quality of assessments are "validity" and "reliability"

- **Validity.** A valid assessment is one that contributes to good decisions by providing accurate information for specific decision purposes. A valid assessment measures what it is intended to measure and permits accurate and fair conclusions about students' performance. To the extent these things are true, an assessment serves its intended purposes and provides a sound basis for decisions. For example, if an assessment is intended to be used to draw conclusions about an individual student's ability to write, the assessment is valid to the extent that it really does allow you to draw that conclusion.

- **Reliability** refers to how repeatable the scores are. Would the conclusions drawn about student ability be the same regardless of when the assessment was administered, which form of the assessment the student received, or which rater happened to score the work? Do two equal scores indicate equivalent performance across students, raters, and days? How comfortable would you feel, for example, if the same student paper received vastly different scores from two different raters, or from the same rater on two different days? Obviously, if an assessment is not reliable, it cannot be valid.

Although the formal terms "reliability" and "validity" are not used very much in this chapter, it's really what this chapter is about—thinking about the reliability and validity concerns of alternative assessment. Familiarity with these issues is important for those either selecting or designing alternative assessments whether teachers or those designing large-scale assessments. Assuring the validity of assessment is easier said than done. It requires consideration of a number of assessment features, which are described below.

(**Activity 5.1**—*How to Be a Good Consumer* parallels each of the topics presented below. **Activity 2.5**—*Clapping Hands* provides a less technical way of considering many of the same issues. **Activity 2.5** may be more appropriate than **Activity 5.1** for teachers and others new to the field. Other activities relevant to specific topics and cross-references to previous chapters will be mentioned as we go. Many of these also provide a less technical approach to the same issues.)

Content and Skill Coverage

The student knowledge and skills covered by an assessment should be important ones, and content should represent enduring themes or significant knowledge, not trivialities. Assessment coverage should closely match the goals of the curriculum, and the assessment device should cover the entire range of that content.

Additionally, not all the goals we have for students are best measured using alternative assessment. Many people believe that the most productive use of alternative assessment takes place when addressing "big" outcomes such as (1) problem solving, reasoning, critical thinking and attitudes; (2) complex applications of knowledge and skills to new settings or situations; or (3) the processes that students use to complete tasks. To assess such complex learning outcomes, many indicators are needed. (See Chapter 2 for further discussion of this issue. **Activity 2.6**—*A Comparison of Multiple-Choice and Alternative Assessment* asks participants to discuss the relative advantages and disadvantages of each approach.)

Discovering what the assessment covers is not as simple as it might first appear. Sometimes, for example, the developer makes a written statement about coverage (e.g., "this assessment measures critical thinking, attitude toward mathematics, and problem solving"), but it is impossible for you to match these statements to what the students are asked to do or what the performance criteria cover. Or, there is no statement about coverage and you have to figure out what the assessment covers from the performance tasks and criteria. Ideally, statements about coverage match what the tasks ask students to do and what is rated on the criteria. Therefore, when we ask about the quality of assessment coverage, we need to consider tasks and criteria in addition to any written statements of coverage made by the developer.

Questions you should ask in evaluating an assessment for appropriate content coverage include:

1. Is it clear what goals, skills and/or content are to be covered on the assessment?

2. Are these goals, objectives or targets best assessed using an alternative assessment?

3. Does the assessment address and embody valued student outcomes and avoid irrelevant or unimportant content and skills?

4. Does the assessment deal with enduring themes or significant knowledge within or across disciplines?

5. Do statements of coverage, task content, and performance criteria match up?

Tasks: General

Every assessment requires that the student complete some task or activity, whether it is a task assigned as part of regular classroom instruction, a task the student sets for himself or herself, or a task developed specifically for the assessment. Tasks can be of many types, such as painting a picture to visually represent the student's view of the atom, answering an open-ended question in mathematics, using manipulatives to add, doing a science lab, explaining the rationale for a problem solution, performing a skit on ionic bonding, developing a computer program that does Mandelbrot and Julia plots, or doing a series of sculptures to represent topological concepts. (Many examples of tasks are provided in Chapter 4 and **Activity 2.10**—*Tasks vs. Performance Criteria*.)

The central question in evaluating tasks is the extent to which the task elicits the included skills or abilities. Did you ever give students an activity that didn't work out as expected? For example, you wanted students to express their understandings of atomic structure by acting it out, but they actually mindlessly repeated words and phrases from the text that they believed you wanted to hear. The point is, we can't assess a skill if the students don't use it. We need to question the degree to which the task elicits the intended performance. For example:

- Does the word problem on the physics of baseball really elicit problem solving?
- Does the *group* task on designing a mechanized vehicle enable one to judge the abilities of *individuals*?
- Will the task on understanding the water cycle result in problem solving or regurgitation of memorized knowledge?
- Does the graphing problem give so many hints that students have no thinking to do; or so few hints that they have no idea what to do?
- Does the task on determining the amount of speeding near the school provide ample time for students to complete the task so that results will reflect actual knowledge or performance, not test-taking skills and memorized algorithms?

In short, do we get what we want?

Since most of the skills assessed using an alternative assessment are fairly complex (e.g., critical thinking, problem solving, communication, and group collaboration), tasks have to be "meaty." And since they are usually complex and time-consuming they should be worth the time and effort of students and teachers—meaningful, worthwhile, and interesting.

Questions for evaluating tasks include:

1. Will the task(s) given to the student likely elicit the desired performances? Do you get what you think you're getting?
2. What is the extent to which the task recreates an "authentic" context for performance? Is the task of importance to the students? Do they find it interesting? Will it elicit "real" vs. "artificial" performance? Will assessment tasks engage and motivate students to perform to the best of their ability?
3. Are the tasks consistent with current educational theory and best practice? Is the task instructionally sound? Does it fit well with instructional goals and practices?

   (Note: In an era of rapid change in educational thought, this can be a troublesome question to address. While it may be easy to identify tasks that are glaring examples of inconsistency with current pedagogy, judgments on many tasks will be quite complex and dependent upon the particular orientation and policy that shape the context in which the assessment will be used.)
4. Have tasks actually been reviewed by content experts to judge content quality and authenticity?

Tasks: Sampling, Representativeness, and Generalizability

Let's say that the goal of an assessment is to make judgments about individual students' ability to problem solve in mathematics. We want to design a test that allows us to draw this conclusion. So, we develop five tasks that we believe require problem solving on the part of the student and randomly distribute them in classrooms; each student only does one of the tasks. Can we use the results of this assessment to draw conclusions about the problem solving ability of individual students? Probably not, because results on any single problem might be the result of a fluke; for example, the student just happened to be given a very similar problem last week in mathematics class.

Could we draw conclusions about the group of students as a whole? (That is; can we average performance together across all the students to see how they are doing as a group?) Well, maybe. It is unlikely that every single performance of *all* students is due to a fluke. Also, any fluke that results in a higher score is fairly likely to be balanced out by a fluke that results in a lower score. The real issue in this case is not so much errors in assessing any individual student but whether the *sample* of five problems adequately *represents* the domain of problem-solving skills and situations in which problem solving is likely to occur. If we give students five other problems, will our conclusions about their problem-solving ability be the same?

This scenario illustrates the issues of sampling, representativeness, and generalizability. Do we have a good enough range of tasks that we can draw the conclusion we want to draw? Would student performance on other possible problems not on the assessment be similar to the performance on the particular set of tasks on the current assessment—in other words, could we *generalize* from the current set of tasks to the whole domain of problem solving?

This is not a trivial issue. There have been many indications that it is hard to predict from performance on one task how students will do on a second task. For example, a student might develop a good experimental design for one problem, but have less success with another problem.

Note that if all we want to do is make a statement about how well students do *on the tasks on the current test*, issues of generalizability are not important. We can always say, "On this test students did thus and so." Generalizability becomes an issue when we want to generalize a conclusion *based on* a test. For example, we might want to say, "Based on this test, we judge that Alex is a competent problem solver." In this case generalizability is of paramount importance. The issue of generalizability is especially important when the conclusions one draws about students result in high-stakes decisions being made—for example, "Sevu has not shown competence in problem solving, therefore, he will not receive a high school diploma."

In the final analysis, issues of generalizability can only be answered by trying out the assessment and actually seeing how well performance generalizes to similar tasks and to other measures of the same skill. However, we should always be thinking about this issue when we select or design alternative assessments. We should ask, "What's my best estimate of whether this set of tasks represents the domain well?" and "What's my best guess as to whether performance on this set of tasks allows me to draw the conclusions I want to draw?"

Of course, the quality of the tasks is not the only thing that influences our ability to generalize. If raters do not score the performances in a consistent manner, then we likewise won't be able to draw conclusions about student ability. This is the issue of interrater reliability—will the same problem solution, experimental design, oral presentation, or group collaboration receive the same score regardless of rater or day?

Important questions about sampling and representativeness incorporate issues of context, scoring, and generalizability:

1. How well do the tasks cover the skill area (domain)? Do they cover what is really important in the area of interest? What evidence exists to show that

performance on these tasks is generalizable to other similar tasks and representative of the desired behavior?

2. Is there a clear statement of the question to be answered (e.g., typical performance or best performance)?

3. Are enough samples of performance or displays of knowledge/skills/behaviors collected for us to say with confidence that results adequately show what a student can do. That is, have we collected a generalizable sample?

4. Is compelling evidence available to show results are consistent across raters and across scoring occasions?

Tasks: Extraneous Interference

Extraneous interference occurs when there is something in the assessment that hinders the ability of students to demonstrate what they know and can do with respect to the skill or knowledge of interest. For example, some performance assessments may require role-playing, even though the ability to play a role has nothing to do with the ability or knowledge being assessed. In this case, the ability to play a role might get in the way of the ability to demonstrate the real skill being assessed, for example, critical thinking. Or, there might be a great deal of reading necessary to set up an assessment task, so much so that the ability to understand the directions interferes with the ability to do the task. An even more subtle type of extraneous interference can occur when the performance depends on *knowledge* that is extraneous to what is being assessed. For example, suppose students are asked to design a dog-run with a specified square footage. In order to design a reasonable run, students need to know that an acceptable run has to be wide enough for a dog (i.e., six inches wide is not adequate) and large enough to provide exercise. Some students may not have this additional information.

(Extraneous interference is also discussed in **Activity 4.1**—*Performance Tasks—Keys to Success.)*

When examining tasks for extraneous influences, look for the following: Are there things (either irrelevant or not critical to the skill actually being assessed) that might get in the way of students being able to demonstrate what they know and can do? Examples of such extraneous factors might include:

- Ability to read
- Ability to write
- Ability to role play
- Ability to understand the context in which activities or tasks do or do not occur

- Personality (shyness, fear of public speaking, etc.)
- Extraneous background information required to do the task
- Physical limitation or condition that interferes with performance (examples: a color-blind student who cannot visually determine the endpoint of a titration; a wheelchair-bound student who cannot carry out appropriate sample techniques in a stream; or a student with speech difficulties who struggles with oral presentations)

Performance Criteria

Performance criteria delineate the characteristics that we will use to judge the quality of a student performance. They are, thus, central to alternative assessment. (**Activity 2.10**—*Tasks vs. Performance Criteria*, provides examples of performance criteria. This activity can be used if participants seem to be unsure as to what performance criteria are.)

If an assessment instrument does not include criteria for judging the quality of student responses to a task, it cannot be considered "assessment." (Even when teachers are engaged in informal assessment—observing, questioning, and discussing with students—they use internal or implicit criteria to infer information about student understandings.) Because notions of quality are so linked to purpose of the assessment and the goals or targets being assessed, evaluating the quality of performance criteria is not simple. Considerations in evaluating the quality of performance criteria include content, clarity, and type.

Content. Content is important for two reasons. First, since performance criteria provide the basis on which student responses will be judged, they should cover *all* the most important dimensions of performance. What sense would it make, for example, to judge the quality of problem solving only by the ability to clearly communicate what was done? We might also want to base our judgments of the quality of a student response on how well the student understood the problem, translated the problem into appropriate pictures and symbols, considered various approaches, went back to check if results made sense, and made adequate calculations.

Secondly, teachers tend to concentrate instruction on what is covered in the performance criteria. Therefore, it is important that criteria deal with all valued dimensions of performance, not just those easy to define and score. For example, consider communication in mathematics. What instructional messages are sent to teachers if we leave out of our criteria the ability to clearly communicate what was done? Since it isn't assessed, teachers might not emphasize it in instruction.

Clarity. Performance criteria give guidance on how to score student responses. How helpful would it be if the criteria only told you to score "problem solving"? Would you know what to look for? Would you know how to differentiate strong problem solving from weak problem solving for third grade students? Would students know what to do in order to do well on the task? Would it be more useful if "problem solving" were carefully defined with samples of student work that illustrated strong, middle, and weak performance?

In order to be useful, criteria must be clearly stated and must be illustrated with samples of student work in such a way as to help teachers and students understand what it takes to be successful.

Type. Criteria may be general or they may be task specific, and they can be either holistic (one score) or analytical trait (more than one score). Chapters 3 and 4 thoroughly discuss these differences and suggest when different approaches might be most useful. To summarize here, generalized analytical-trait criteria might be best used in the classroom or when the skills being assessed are the "big" ones (such as problem solving, critical thinking, group collaboration, and science process). Task-specific criteria are frequently used in large-scale assessment because training raters is quicker and scores are more reliable.

(**Activity 2.9**—*Going to School* and **Activity 4.3**—*Performance Criteria—Keys to Success* provide examples of different types of performance criteria and their relative advantages and disadvantages. **Activity 3.1**—*Camping Trip,* **Activity 3.2**—*Sorting Fish,* and **Activity 6.2**—*How Can We Know They're Learning?* illustrate how developing generalized performance criteria help us to more clearly understand the goals we have for students. **Activity 2.5**—*Clapping Hands* illustrates the need to communicate criteria to students so that they will know what it takes to succeed.)

Questions to ask about performance criteria include:

1. Do the criteria match the developer's stated "coverage" of the assessment? For example, if a developer claims that an assessment assesses mathematics problem solving, ability to communicate mathematically, and group cooperation, are there criteria for judging all these things? There should be criteria for judging everything claimed to be assessed.

2. Do the criteria cover what is important? Are the concepts, processes, skills, etc. that you wish to measure appropriately addressed by the criteria? For example, if you want to assess problem solving, are all the important dimensions of a good problem solution covered?

3. Are the criteria stated clearly enough so that raters can agree on meanings? Are there examples of student work to illustrate the criteria?

Additional questions to ask if the performance criteria are intended to be useful instructional tools are:

biases in assessment tasks and developing ways to overcome such biases is a difficult but critical part of designing and evaluating assessments, whether alternative or traditional.

(Fairness and equity issues are explored in **Activity 5.2**—*Dr. Know-It-All*, **Activity 5.3**—*Chickens and Pigs*, **Activity 5.4**—*Equity Role Play*, and **Activity 5.5**—*Questions About Culture and Assessment*. These issues also frequently arise during **Activity 2.5**—*Clapping Hands* and **Activity 2.7**—*Think of a Time*.)

Important fairness questions include:

1. Do the content or context of the task or assumptions about expected performances reflect knowledge, experiences and values that are equally familiar, acceptable, and appropriate for all groups of students?
2. Does the assessment tap knowledge and skills all students have had adequate opportunity to acquire?
3. Is the assessment as free as possible of cultural, ethnic, or gender stereotypes?

Important rater bias issues include:

1. Is some feature of the performance (e.g., handwriting, poor spelling, pet peeves) influencing how another, supposedly independent feature is judged?
2. Does knowledge of the type of student performing the task (e.g., gender, ethnic heritage, curriculum track) influence judgments?
3. Does knowledge of individual students affect judgment of performance? (e.g., "Susie always does such good work, I'll give her the benefit of the doubt. Heather always does such poor work, someone else must have done this for her.")

Consequences

Any assessment may have unanticipated side effects. A positive unanticipated side effect might be an assessment that improves a student's attitude toward a subject area because the assessment was fun to do. Negative effects might include restricting curricula to what can be easily assessed; communicating unintended messages about power, control, and social roles/status in the classroom; or providing narrow and restrictive images of the nature of knowledge in a particular subject area.

The history of testing is replete with examples of good intentions gone awry: tests that were supposed to improve the educational system but which actually resulted in narrowing instruction to basic skills; or multiple-choice tests, originally developed to provide more "objective" measures of achievement that actually resulted in perpetuating the idea that there is always a right answer to any problem. Good intentions are clearly not enough. Cognizant of past problems, we need to design assessments that minimize unintended negative consequences.

Science and mathematics education have been particularly plagued by several myths that have been reinforced by traditional assessments. These include:

- One right answer exists to all mathematics and science problems.
- Computation is central to mathematics and science.
- "Knowing" is more important than being able to do (for example, being able to pick a correct answer is more important than being able to carry out a lab activity).
- Science and mathematics are primarily done with pencil and paper.
- Being able to recognize technical terms is more important than understanding concepts.
- Some external authority is the only valid judge of whether an idea in science or mathematics is acceptable or "true."

These beliefs have a powerful influence on student learning, use of knowledge, interest in and attitude toward subject areas, and access to knowledge in particular areas. Assessment that communicates such notions can have a strong negative impact on student lives. This does not mean that factual knowledge is unimportant. Emphasizing reasoning without concepts and facts is impossible. Clearly, a balance is needed.

(Discussion of how assessment communicates what is valued can be found in Chapter 2. Discussion of unintended consequences of assessment often arise as part of **Activity 2.5**—*Clapping Hands* [effects on students], **Activity 2.7**—*Think of a Time* [effects on students], **Activity 2.9**—*Going to School* [effects on instruction], and **Activity 3.1**—*Camping Trip* [effects on instruction].)

When examining the consequences of an assessment, consider the following questions:

1. What will this assessment communicate to students, teachers, and others about:
   - What is important to teach? Will there be effects on instruction and curriculum?
   - What is important to learn? For example, a student might say, "I don't have to worry about problem solving because all the problems on the test come from homework."

- What is the nature of the discipline(s) being assessed, and the nature of knowledge in those subject areas? For example, is an assessment sending a message that mathematics is primarily computation?
- Who has the authority to judge what is acceptable knowledge and performance in these areas?
- What images of learning, and of life, are being communicated?
- How should we teach? Does the assessment exemplify good instruction?

2. What effects might the assessment have upon the lives, attitudes, emotions, self-image, career options, etc. of the teachers, students, and parents involved?
3. Is the assessment worth the instructional time devoted to it? Will students learn something from the assessment itself?
4. Does the assessment provide information that is relevant to the decisions being made with that assessment?
5. Will the assessment be perceived by students and teachers as a valid indicator of student competence in the particular area being assessed?

Cost and Efficiency

An assessment should be practical. The consequences of the assessment should be worth the cost and effort required to obtain it. By necessity, assessment design is always a series of trade-offs. For example, observing students doing laboratory activities may provide a more accurate picture of student lab skills than examining laboratory notebooks. However, examining notebooks is more efficient. Does the increase in accuracy justify the increase in time needed for direct observation? The issue is whether the loss in validity is acceptable when compared to the increase in efficiency.

Impacts and effects can include more than just obtaining good data. Involving students in rich, engaging tasks and in self-assessment can promote the very learning that is being assessed. Also, involving teachers in the design and scoring of alternative assessments can be an extremely powerful professional development experience. This makes the additional cost of alternative assessment more worthwhile.

In evaluating assessment design, it is important to identify the relevant trade-offs and determine the acceptability of gains or losses in validity, instructional usefulness, reliability, and efficiency. Although addition of authentic, alternative, and performance-based tasks usually raises the cost of an assessment, an assessment that meets the criteria described in this chapter will more likely be worth the investment. Questions that must be considered are:

1. Is the assessment cost-efficient? Are the effort and the resources required to do the assessment worth the results and consequences?

2. Is the assessment "practical" or "do-able"? (For example, does the assessment require so much time, equipment, space, etc. as to preclude use? Are there safety or privacy concerns?)

Conclusion

Alternative assessment can be a powerful tool for learning if we do it correctly. We hope the considerations and activities in this chapter will help us all to be wise consumers and good developers of alternative assessments.

Further Readings

Brandt, R. (1992). On performance assessment: A conversation with Grant Wiggins. *Educational Leadership, 49*(8), pp. 35-37.

Dunbar, S., et al. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education, 4*(4), pp. 289-303.

Fitzpatrick, R., & Morrison, E. (1971). Performance and product evaluation. In R. Thorndike (Ed.), *Educational Measurement.* Washington, DC: American Council on Education.

Johnson, S., & Bell, J. (1985, Summer). Evaluating and predicting survey efficiency using generalizability theory. *Journal of Educational Measurement, 22*(2), pp. 107-119.

Linn, R., et al. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), pp. 15-21.

Popham, J. (1993, February). Circumventing the high costs of authentic assessment. *Phi Delta Kappan, 74*(6), pp. 470-473.

Rock, D., & Nelson, J. (1992, Summer). Applications and extensions of NAEP concepts and technology. *Journal of Educational Statistics, 17*(2), pp. 219-232.

Shavelson, R., & Baxter, G. (1992, May). What we've learned about assessing hands-on science. *Educational Leadership, 49*(8), pp. 20-25.

Shavelson, R., et al. (1990, May). Indicators of science achievement: Options for a powerful policy instrument. *Phi Delta Kappan, 71*(9), pp. 692-697.

Shavelson, R., et al. (1991). Steps in designing an indicator system. *ERIC/TM Digest*. Washington, DC: ERIC Clearinghouse on Tests, Measurement and Evaluation.

Shavelson, R., et al. (1991). Performance assessment in science. *Applied Measurement in Education. 4*(4), pp. 347-362.

Shavelson, R. , et al. (1992). Performance assessments: Political rhetoric and measurement reality. *Educational Researcher, 21*(4), pp. 22-27.

ACTIVITIES

Activity 5.1

How to Be a Good Consumer of Science/Mathematics Alternative Assessments

- or -

How to Tell the Gems from the Duds

Purposes:

1. To practice making judgments about the quality of science and mathematics alternative assessments
2. To develop expertise on the characteristics of good assessment
3. To prepare participants for developing their own assessments by critiquing some developed by others

Rationale:

Contrary to popular belief, alternative assessment does not automatically mean good assessment. Moving to alternative assessments will not automatically solve all our assessment (and political) problems. The goal of this activity is to help participants become good consumers of alternative assessments in science and math. Specifically, this activity discusses criteria for what constitutes good alternative assessment in science and mathematics and has participants apply this knowledge by reviewing and critiquing actual performance assessments. This knowledge will also help participants design their own assessments.

While good quality assessment is important in any setting (classroom to large-scale), the topics dealt with below take on special importance when alternative assessments are to be designed for "high stakes" purposes such as high school graduation.

This activity presents a fairly technical and thorough examination of technical issues. It is recommended for those who already have some experience with alternative assessment. Beginners should use other activities to explore the notion of technical quality (these are noted both in the chapter text and below). An especially good beginning activity for classroom teachers is **Activity 2.5**—*Clapping Hands.*

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- Transparencies made from overheads at the end of this chapter
- One copy of Chapters 2, 3, 4, and 5 for each participant
- **Handout A5.1,H1**—*Alternative Assessment Evaluation Form*
- Sample assessments from **Appendix A** (you will only need one copy for every three participants; however, participants often want a complete set to take home;) choose the sample assessments to illustrate the specific technical issues you wish to emphasize; assessments that illustrate various points are listed in each section below

Time Required:

6 hours

Facilitator's Notes:

The general procedure is to go over criteria for quality one section at a time with illustrations, and then have participants work in small groups to critique sample alternative assessments using the same criteria. The critique form is **A5.1,H1**—*Alternative Assessment Evaluation Form*. Sample assessments are provided in **Appendix A**. Notes in each section suggest assessments to use. (Note: Suggestions on uses of samples are based on the material included in the *Toolkit*. Be sure to let participants know that newer versions of the assessments, or other portions of the same assessment, may look different. For example, samples of student work might not be included here, but might be available from the author.)

In the steps below, sample statements to participants are in italics.

A. Introduction

1. Before beginning the actual activity, do the usual openings: Do introductions. Go over the purposes for the activity and the handouts. Ask participants if they have any particular questions that they want discussed, etc. When this is done, introduce the activity focus:

Contrary to popular belief, alternative assessment does not automatically mean good assessment. Alternative assessment will not automatically solve all our assessment problems. Use **Overhead A5.1,O1** to illustrate this point.

The purpose of this component of the training is to help participants be good consumers of alternative assessment devices, and to know what to do when they are designing performance assessments. The things that we will be looking at in our sample alternative assessments include coverage, performance tasks, performance criteria, bias and fairness, consequences, and cost efficiency. Use **Overhead A5.1,O2**. Briefly go over the categories and show where to find the relevant sections in the Chapter 3, 4, and 5 handouts. **Handout A5.1,H1** is a summary of what to look for.

2. Mention that performance assessments have three major design components: performance criteria, performance tasks, and context. The task is what you give students to do—for example, write a paper, do a lab, answer a question. The performance criteria are used to judge performance on the task. The context describes such things as the purpose for the assessment, grade levels, and reporting requirements. We will be looking at these components separately. (Chapter 4 provides lots of examples of tasks and performance criteria. **Activity 2.10**—*Tasks vs. Performance Criteria* specifically delineates the task and performance criteria parts of an alternative assessment.)

B. Content and Skill Coverage

1. **Overhead A5.1,O3** summarizes what to look for.

   What does the developer say is covered by the assessment? Some possibilities are: facts, knowledge, application of concepts, science process skills, group process, critical thinking skills, reasoning, problem-solving ability, etc.

2. Participants should try to find a statement by the author of the skills to be assessed. (Clear statements are in **Samples 4.1, 4.2, 4.3, 4.5, 4.6, 4.7, 5.2, 5.5, 5.6,** and **5.7**. Unclear statements of coverage are in **Samples 4.4, 4.9,** and **5.4**.) If there is no statement, participants must determine coverage from a description of the tasks or the performance criteria. You have to infer coverage from performance criteria in **Samples 5.3** and **5.9**.

3. Then participants need to consider whether alternative assessment is appropriate for this content. Alternative assessments are most useful for outcomes such as: application of concepts or skills; "big" outcomes such as problem solving, collaborative working, and critical thinking; or "real time" performances that have to be observed, like proper use of laboratory equipment. Alternative assessments may not be the most appropriate method of assessment for all types of content and all assessment contexts. (**Samples 4.5, 4.9,** and **5.8** are examples that provoke a thoughtful discussion of this issue.) Chapter 2

presents additional information on the types of student skills best assessed using an alternative assessment.

C. Performance Tasks

Tasks include such things as a math problem to solve, a science lab to do, a group project, or journal writing. Different types of tasks are discussed in Chapter 4. **Activity 2.10**—*Tasks vs. Performance Criteria* helps participants identify the tasks in an alternative assessment. The strategy for this section is to first show participants different types of tasks for science and mathematics assessment, go over some things they should consider when critiquing these tasks, and then give them practice reviewing tasks. **Activity 4.1**—*Performance Tasks—Keys to Success* can be substituted for this portion of **Activity 5.1**. **Activity 4.2**—*Spectrum of Assessment Tasks* and **Activity 2.5**—*Clapping Hands* can be used to supplement this section.

1. **Characterize Tasks. Overhead A5.1,O4** outlines dimensions that seem to characterize current science and math alternative assessment tasks. These dimensions of tasks are described in Chapter 4, with examples. You can use the samples in Chapter 4 to make overheads. The dimensions are:

    a. Open-response vs. open-ended

    b. Written vs. hands-on activities

    c. Type of response—written, oral or visual

    d. Length or complexity of the task—on-demand performance assessment, project or portfolio

    e. Amount of group work

    Show several examples of each type, and then ask the participants to describe one or two additional tasks using these dimensions. The goal here is to get participants to attend to differences in tasks so that they can begin to identify the approaches they like best. You can even have participants begin to articulate what they like and dislike about the examples they are seeing.

    It is not necessarily the case that any single type of task is better than any other. Participants need to design/choose tasks that correspond to their purposes, grade levels, student sophistication, and philosophy of instruction.

2. **Evaluate Tasks.** What makes a good task? The "task" section in Chapters 4 and 5 and **Overheads A5.1,O5** through **A5.1,O7** list several things to look for:

General:

a. *Will the task elicit the desired performance? The task has to match the stated "coverage" of the assessment. For example, if the developers say that the assessment measures problem solving, will the task actually result in problem solving being done? As an illustration, the goal of Vermont's portfolio system is to assess student problem solving in math. Although some guidance was given as to the sorts of things students could put in their portfolios, students were pretty much on their own regarding what they selected to show their abilities as problem solvers. The first year, over half of the material in the portfolios were worksheets. It was impossible to assess problem solving because students did not submit performances that demonstrated problem solving. The task did not elicit the behavior that they wanted to assess.*

It seems almost silly to say it, but if you want to assess science methodology, ability to problem solve, critical thinking, or group process skills, you need to have a task that results in students demonstrating these skills. It's amazing how easy it is to design a task that doesn't elicit the desired skill. Sometimes the only way to know for sure is to try it out.

b. *Is the task "authentic" enough? What would a real-life task in math or science look like? You can have the participants brainstorm this. It might be a self-generated problem or task that needs to be solved, a self-generated purpose for the task, a real audience, etc.*

Obviously, it is impossible to follow students around until they spontaneously exhibit the behavior of interest. By necessity, then, we have to build tasks that are not completely like real-life. Use **Overhead A5.1, O8** to further demonstrate this point.

The question then, is not "is it authentic?" but is it "real enough" to elicit the desired behavior on the part of the student?

c. Does the task reflect current educational theory and best practice? The importance of this is covered in Chapter 3. Remember assessment tasks are used by teachers for models of instruction.

d. Developers also have the responsibility to have their products reviewed by outside experts. Is there any evidence that tasks were reviewed?

Samples that can be used to illustrate the "authenticity" or "meatiness" of tasks are:

- Low meatiness: **Samples 4.5** and **5.8** (some people think these look like multiple choice with the choices taken off)

- Medium meatiness: **Samples 4.8** (all written, only one right answer), **4.3** (hands-on, but lock-step), **4.6** (all written but more open-ended)

- High meatiness: **Samples 4.1, 4.10, 5.1, 5.3, 5.9**

Sampling:

Any time you collect information about a student's performance, you are "sampling" that behavior because it is impossible to gather every instance of a behavior. The question is whether the sample you collect really does represent the student ability in question—see **Overhead A5.1,O6**.

a. Are enough samples collected to make a judgment about individual students? Many performance assessments fail on this point. Because they are so time-consuming to give, students are usually only asked to do one or two different tasks. How much can you generalize about student problem solving, for example, from one or two open-response math problems?

b. Do the performances cover the ground? We can't generalize about students' ability to write in general, if all we've tested is their ability to write stories. What about their ability to write up labs or explain what they did on a math problem? In order to make a general statement about student ability, one that goes beyond the task at hand, we have to have different performances that together cover the area of interest. For example, if we want to see whether students can utilize a variety of problem-solving strategies, we need to devise tasks that require a variety of strategies.

Samples that can be used to illustrate sampling:

- Students only get one task: **Samples 4.2, 4.4, 5.1, 5.10**

- Several tasks are available: **Samples 4.3, 4.7, 5.2**

Extraneous Interference:

Extraneous interference occurs when there is something in the assessment, having nothing to do with the skill being assessed, that gets in the way of the students being able to demonstrate what they know and can do. For example, some performance assessments require role-playing, even though the ability to play a role has nothing to do with the skill being assessed. An example is science/technology/society exercises where students explore a *current social issue by bringing scientific information to bear. Students play the parts of different players in the controversy.*

Other examples of possible sources of interference are listed in the handouts and on **Overhead A5.1,O7**. Remember that it is only interference if the task requires a skill that is irrelevant to the skill being assessed. Examples of possibly extraneous features are:

- Ability to read (for example, on a problem-solving test)
- Ability to write (for example, on a math test)
- Ability to role-play
- Personality (for example, shyness, fear of public speaking)

Samples that can be used to illustrate extraneous interference are:

- **Sample 4.2** (lots of writing)
- **Sample 4.8** (lots of reading)
- **Sample 5.4** (group work)

3. **Practice.** Pass out the sample math and science performance assessments from Appendix A that you have chosen for participants to use. The participants should work in small groups to do the following. (Use the evaluation form as a worksheet.)

   a. Characterize the task(s) in the performance assessment using the categories in #1 above.

   b. Discuss whether they feel the task will elicit the desired behavior.

   c. Discuss whether they feel the task is "authentic" enough.

   d. Discuss the extent to which they feel the task represents best practice.

   e. See if they can find evidence that the tasks were reviewed by outside experts for the characteristics in b-d.

   f. Review the task(s) for issues of representativeness and generalizability.

   g. Review the task(s) for extraneous factors that might interfere with a student's performance.

   Get the group back together to discuss what they found.

D. Performance Criteria

Chapter 3 discusses the importance of designing our performance assessments to be useful instructional tools. Chapter 4 looks at different types of performance criteria and their relative advantages and disadvantages. The strategy for this section is to describe the types of performance criteria we are seeing in science and math, review some criteria for performance criteria, and then have participants practice applying the criteria to samples.

Activity 4.3—*Performance Criteria—Keys to Success* can be substituted for this portion of the activity. **Activity 2.5**—*Clapping Hands* can supplement the discussion.

1. **Characterize Criteria. Overhead A5.1,O9** lists types of performance criteria that are being used to score performance assessments. These are described in detail in Chapter 4, with examples. Use these examples to make overheads. (**Activity 2.9**—*Going to School* also can be used to illustrate different kinds of performance criteria.) The types of criteria are:

   a. Task specific vs. generalized

   b. Holistic vs. analytical trait

   c. Quality vs. quantity

   Show examples of each type of scoring, and then ask participants to characterize one or two more. The goal here is to get participants to recognize the various scoring mechanisms when they see them.

2. **Criteria for Criteria.** The guidelines for reviewing performance criteria are described in Chapter 5. They are also listed on **Overhead A5.1,O10**.

   a. *Do the performance criteria relate to what is purportedly covered by the assessment? For example, if the developer claims that the assessment covers (or assesses) problem solving, group process skills, concepts related to physics, and self-reflection, are there criteria to rate each of these? Even if the task requires these features, the assessment cannot be said to "assess" these things if there are no performance criteria with which to make judgments concerning them.* Samples that match-up well are **4.7** and **5.1**. Samples that might not match up well are **4.5** and **5.8**.

   b. *Do the criteria cover what is important? For example, in the Informal Writing Inventory, a screening instrument for special education, one portion requires students to look at a picture and write a paragraph telling about what is happening in the picture.* A description of the task and the scoring procedure is provided on **Overhead A5.1, O11**.

Now, granted this assessment might just focus on grammar, but you would probably not want to draw any conclusions about student ability to write from just looking at grammar. Thus, the criteria by which this performance was evaluated do not cover all the relevant dimensions of the task.

Although this is an extreme case, some large-scale performance assessment systems tend to go in the same direction. For example, the Illinois analytical trait writing assessment does not score voice and word choice because the developers felt that it would be too difficult to get consistency in scores; it is just too personal. What does this communicate to teachers about what is important to include in instruction? Anecdotal reports from teachers in Illinois indicate that, in fact, teachers do concentrate their instruction on the traits measured in the assessment. Do we leave things out just because they are difficult to define?

(Return to **Overhead A5.1,O10**)

c. *Are the criteria stated clearly? Are they illustrated with samples of student work? The performance criteria essentially define the goal you have for students. In order to be useful tools, they need to be defined so carefully that you can get consistency in scoring. Such care in definition is what makes them so useful. No more fuzzy, global statements. Performance criteria should have score points defined and should have samples of student work to illustrate what the scores look like. The samples are also used to help students understand what it is they are supposed to do.*

- Samples with sketchy criteria are **4.3** and **5.9**
- Samples with some description are **4.7**, **5.1**, **5.2**, **5.4**, and **5.5**
- Samples with more description are **4.2** and **4.6**

d. If the assessment is intended to be useful as an instructional tool, then the performance criteria must also be general, be easily taught to students, and cover broad concepts. Are the criteria general, or are they task specific? Examples of criteria are given in Chapter 4.

We would like to propose that more generalized scoring (for "big" outcomes) is more useful as an instructional tool. When criteria are tied to tasks, they do not generalize across tasks. That is, the criteria enable you to score this task, but not the next; you can't generalize them. We should be aiming for performance criteria that provide an overall picture of the target, not just how the target manifests itself in a single problem. With generalized criteria, such as Oregon's, evaluating one problem solution will help you evaluate the next solution because the goal is to understand what good problem solving is in general; the goal is generalization.

Of course, in order to use the system well, raters (including teachers and students) have to see many samples of student work. Thus, it takes longer to train raters than when using systems where criteria are tied to tasks, and consistency of judgments between individuals, at least at first, is lower. That's why criteria tied to tasks are used in many large-scale assessments. It is easier and faster to train raters, and reliabilities are high right from the start. However, in the long run we, and students, will have greater understanding if we aim toward more generalized performance criteria. (Which would you rather have your students do: score a bunch of performance tasks where each task used different criteria, or use a more generalized evaluation procedure that tried to analyze what makes problem solving, for example, good in general?)

If you accept the premise that it is necessary to build performance assessments that are instructionally useful, then you should develop generalized scoring criteria for "big" outcomes such as problem solving, critical thinking, group process skills, and communication.

e. *Can the criteria be easily taught to students? Is it useful to teach these criteria to students in order to enhance self-assessment and promote learning? This once again touches on instructional usefulness. One thing to think about is memory load. Too many criteria are hard to remember. Too few might leave something out. Another thing to think about is whether you would want to teach the criteria to students. Would these criteria be useful to enhance self-assessment?*

3. **Practice.** Have participants refer once again to the evaluation form and sample performance assessments. Participants should:

   a. Describe the type of performance criteria used in the assessment.

   b. Evaluate the criteria for clarity, presence of samples of student work, whether they cover the desired "stuff," whether they are task specific or general, and whether they could/should be taught to students.

   Get the group back together to discuss what they found.

E. Bias/Fairness

1. Two issues are discussed here: fairness in the tasks and rater bias. Fairness of tasks really is an issue related to tasks, but is of such importance that we have listed it separately. All students should have equal opportunity to demonstrate their knowledge and skills. Sometimes tasks favor one group more than another. Some examples are provided in Chapter 5.

   Rater bias occurs when ratings are based on something other than the skill of interest. In other words, there is something within raters that affects their *objectivity.* ***Many times, these things are unconscious.*** *For example, raters might let handwriting affect how they score an open-ended problem in math. Or, knowledge of the gender or race of the student might affect judgments of competence. (For example, "That wasn't bad, for a girl.")*

Examples of potentially biasing factors are listed in Chapter 5 and on **Overhead A5.1,O12**.

2. Have participants review their samples (using the evaluation form) for possible sources of unfairness or rater bias and discuss what they find.

F. Consequences

1. *Assessments can have unanticipated side effects. Negative examples include restricting the curriculum to what is on the test (tests rarely cover everything of interest), promoting multiple-choice thinking, and communicating who has the power and knowledge to judge. A positive example would be an assessment that improves a student's attitude toward a subject area because the assessment was fun to do. Sometimes the positive side effects outweigh questions about the technical adequacy of the assessment.*

   A list of things to think about is included in Chapter 5 and on **Overhead A5.1,O13**.

2. Have participants review their samples (using the evaluation form) for possible positive and negative consequences and discuss what they find.

G. Cost and Efficiency

An assessment should be practical. Alternative assessments are generally very costly and time-consuming. The real question is whether the advantages of the alternative assessment outweigh the extra time and expense. We maintain that if the assessment is designed to fulfill the instructional needs of teachers, then it will be cost effective. If not, then it might not be. Issues are listed on **Overhead A5.1,O14**.

Have participants review the assessments for cost effectiveness (using the evaluation form) and discuss what they find.

H. Closing

Ask participants if they have any final thoughts. Reiterate the major points of today's session (see **Overhead A5.1,O15**):

1. Alternative assessment doesn't automatically imply better assessment.

2. Alternative assessments have both advantages and disadvantages.

3. To realize their advantages, you have to use/design alternative assessments carefully.

4. We recommend always using/designing alternative assessments to be useful instructional tools. This will increase their cost effectiveness.

Alternative Assessment Evaluation Form

	Yes	Somewhat	No
1. Content and Skill Coverage	3	2	1

- Clear goals, skills, content to be covered
- Alternative assessment appropriate to measure these goals
- Avoids irrelevant and/or unimportant content
- Deals with enduring themes or significant knowledge

2. Performance Criteria	3	2	1

- Performance criteria match coverage and task
- Performance criteria include everything of importance and omit irrelevant features of work
- Performance criteria are stated clearly and with detail
- There are examples of student work to illustrate the performance criteria
- Performance criteria are stated generally, especially if the intent is use as an instructional tool
- Performance criteria promote a clearer understanding of the skill being assessed

3. Tasks	3	2	1

General:

- Elicit the desired performances or work; match the performance criteria
- Recreate an "authentic" context for performance
- Exemplify good instruction
- Are reviewed by experts
- Are themselves episodes of learning

Sampling/Representativeness/Generalizability:

- Sample of performance is representative of what a student can do
- Domain covered well, generalizable

Extraneous Interference:

- Absence of factors that might get in the way of students' abilities to demonstrate what they know and can do

	Yes	Somewhat	No
4. Fairness and Rater Bias	3	2	1

- Features of performance do not influence how other, supposedly independent, features are judged
- Knowledge of the type of student does not influence judgments
- Knowledge of individual students does not affect judgments
- Task content and context are equally familiar, acceptable, and appropriate for all students in group
- Assessment taps knowledge and skills all students have had adequate time to acquire in class
- As free as possible of cultural, ethnic, or gender stereotypes

5. Consequences	3	2	1

- Communicates appropriate messages
- Acceptable effects on students, teachers, and others
- Worth the instructional time devoted to it
- Provides information relevant to the decisions being made
- Will be perceived by students and teachers as valid
- Students learn something from doing the assessment and/or using the performance criteria

6. Cost and Efficiency	3	2	1

- Cost efficient—the results are worth the investment
- Practical/"do-able"

A5.1,H1

TO AN ASTOUNDING NUMBER... THE DEVELOPMENT OF PORTFOLIO ASSESSMENT SEEMS TO CARRY THE SAME KIND OF SYMBOLISM AS THE SEARCH FOR THE HOLY GRAIL, THAT ELUSIVE SOMETHING THAT, WHEN WE FIND IT, WILL SOLVE ALL OF OUR PROBLEMS.

Joanne Eresh
Pittsburgh Public Schools
Portfolios Assessment as a
Means of Self-Directed Learning
April 1990

A5.1,O1

LABORATORY
NETWORK PROGRAM

Pitfalls

- Content and Coverage
- Performance Criteria
- Tasks
- Fairness/Bias
- Consequences
- Cost and Efficiency

A5.1,O2

LABORATORY
NETWORK PROGRAM

Content and Skill Coverage

- Assessment target clear?
- Is this the most cost-effective way to assess the target?
- Avoids irrelevant and/or unimportant content?
- Emphasizes enduring themes or significant knowledge?

A5.1,O3

LABORATORY
NETWORK PROGRAM

Dimensions of Tasks

- Open-response vs. open-ended
- Written vs. hands-on activity
- Type of response—written, oral, or visual
- Length/complexity—on-demand performance assessment, project, or portfolio
- Amount of group work

A5.1,O4

LABORATORY
NETWORK PROGRAM

Characteristics of Tasks: General

- Elicit right performance?
- Authentic enough?
- Consistent with educational theory?
- Reviewed by outside experts?

A5.1,O5

LABORATORY
NETWORK PROGRAM

Characteristics of Tasks: Sampling

- Enough performances?
- Represent domain?
- Generalizable?

A5.1,O6

LABORATORY
NETWORK PROGRAM

Characteristics of tasks: extraneous interference

What can get in the way of accurate measurement?

- Reading
- Writing
- Role playing
- Personality
- Other?

A5.1,O7

LABORATORY
NETWORK PROGRAM

Realism Is Relative

"...there are many degrees and kinds of artificialities in tests....Performance and product evaluation is one in which some criterion situation is simulated to a much greater degree than is represented by the usual paper and pencil test....There is no absolute distinction between performance tests and other classes of tests—the performance test is one that is relatively realistic."

Fitzpatrick & Morrison
Educational Measurement, 2nd Ed.,
1971

A5.1,O8

LABORATORY
NETWORK PROGRAM

Types of Performance Criteria

- **Task specific vs. generalized**
- **Holistic vs. analytical trait**
- **Judgments based on "quality vs. quantity"**

A5.1,O9

LABORATORY
NETWORK PROGRAM

Criteria for Criteria

- Match coverage
- Cover what's important
- Stated clearly
- General
- Teachable to students

A5.1,O10

LABORATORY
NETWORK PROGRAM

Informal Writing Inventory

The Informal Writing Inventory is a diagnostic procedure.... The results of analyzing samples are brought together into categories that simplify a judgment about the presence, degree and, to a limited extent, the cause of writing disability.

To motivate students to write, 14 illustrations are included.

The identification of writing disability depends on two relationships:

1. The *error index,* which is the ratio of erroneous writing to well-formed writing
2. The *communication index,* which is percentage of total errors that disrupts communication

Informal Writing Inventory
Scholastic Testing Service
1986

A5.1,O11

LABORATORY
NETWORK PROGRAM

Fairness/Bias

Rater Bias

- Pet peeves
- Stereotypes
- Knowledge of individual students

Fairness

- Task context equally familiar and appropriate
- Equal opportunity to learn
- Free from stereotypes
- Language

A5.1,O12

LABORATORY
NETWORK PROGRAM

Consequences

What messages am I sending with respect to:

- Content/what's important
- Control
- How the world is.
- The nature of science and math
- What constitutes good instruction

What effect will this have on:

- Students
- Teachers
- Others
- Curriculum and instruction

A5.1,O13

LABORATORY NETWORK PROGRAM

Cost and Efficiency

- Cost effective?
- Do-able?

A5.1,O14

LABORATORY
NETWORK PROGRAM

Final Thoughts

- Alternative assessment is not automatically better assessment
- Alternative assessments have advantages and disadvantages
- Take a balanced approach to assessment
- Design alternative assessments to be tools for learning

A5.1,O15

LABORATORY
NETWORK PROGRAM

Activity 5.2

Dr. Know-It-All

Purposes:

1. To provide a nonthreatening/humorous opening context for discussing issues of equity and bias in assessment

2. To provide a stimulus for discussing hidden consequences of assessment

Rationale:

Humor can assist in establishing a comfortable space for discussing potentially uncomfortable issues such as equity and bias in instruction and assessment. This video skit satirizes many of the common myths about the nature of science and the people who do it.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- VCR and monitor
- Video clip of "The World of Dr. Know-It-All. Today's Topic: Air Pressure All Around Us," a *Saturday Night Live* sketch starring Dana Carvey, Demi Moore and Phil Hartman (circa 1988-1990)

Time Required:

Video length: approximately 5 minutes

Activity length: approximately 20 minutes

Facilitator's Notes:

1. Introduce the video only by saying something like: *I'm going to show you a video clip of a teaching/learning situation. As you are watching, take mental note of the messages being sent by the instruction/assessment activities.*

2. It may be several seconds before some participants realize that the video is a comedy sketch, particularly if they are new to American culture. If some participants appear confused after a minute or so, note that the segment is from a TV comedy show.

3. As the skit progresses, you may wish to point out some of the more subtle messages being sent (for example treatment/value of animals, which 'child' gets to manipulate the equipment during the activities, the extremely directive prompting in questioning, looking for one right answer, etc.). Messages such as the role of girls in science, science as authoritarian, science as received knowledge/fact, science as a domain for experts, etc. are made very explicit in the sketch.

4. After the video, ask participants to discuss the biases, stereotypes, and implicit messages they observed in the video. Then, ask them if they can relate any instances from their school experiences where they received implicit or explicit messages through assessment activities. You may wish to provide an example or two from Chapter 5 or your own experiences to frame the discussion. This discussion can be used as a springboard to explore equity issues, lead into further equity activities, or lead to Chapter 5's assessment rating activities.

Activity 5.3

Chickens and Pigs: Language and Assessment

Purposes:

1. To increase awareness of how language and experience can influence assessment results
2. To bring to the surface additional equity issues and considerations when designing, using, and interpreting assessments

Uses:

1. In Chapter 2 to illustrate the need to examine more closely the sometimes subtle biases in assessment that can mask children's true learning
2. In Chapter 5 as an opener for a much deeper examination of the interaction of culture, language, and assessment
3. Related activities:
    - **Activity 2.7**—*Think of a Time*
    - **Activity 5.1**—*How to Be a Good Consumer of Alternative Assessments*
    - **Activity 5.4**—*Equity Role Play*
    - **Activity 6.2**—*How Can We Know They're Learning?*

Rationale:

This awareness activity emphasizes the critical role that language and culture plays in effective and equitable assessments. It calls attention to a sometimes hidden dimension of equity in assessment—different meanings for the same term when interpreted in English and another language. Participants consider a common mathematics assessment task and generate their own ideas about language issues that arise when an assessment is used across languages and cultures. It is intended to prompt individual and small group reflection of this and similar issues.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens
- **Overhead A5.3,O1**—*Chickens and Pigs: Language and Assessment* (this can also be used as a handout)

Time Required:

10-20 minutes

Facilitator's Notes:

1. Tell a story from your own experience (or use one of the stories below) regarding an equity issue that was hidden or difficult to recognize. Ask participants to consider a possible problem in the task that might prevent teachers from getting an accurate and rich picture of students' learning, or which limits children's ability to portray their learning well. Here are three possible stories to get things started:

   a. Briefly introduce the "Chickens and Pigs" task (**A5.3,H1**—*Chickens and Pigs: Language and Assessment* and **A5.3,O1**—*Chickens and Pigs: Language and Assessment*) and identify it as a sample from early NCTM work designed to help teachers use a standards-based rubric to score student problem-solving responses.

   Show the task and call attention to the specific language problem that exists for some children and teachers. The following story from professional development work with teachers in Ulithi Atoll in the state of Yap, Federated States of Micronesia, can be used:

   Hoping to help teachers clarify what they value about student problem solving, and then build criteria from the key qualities of fine work identified in their small group discussions, the Chickens and Pigs task was given to all. Teachers began examining the task and some sample student responses. Discussing the task in their own language, puzzled faces turned to the facilitator and a senior member of the group asked if she realized that the word for legs in their language included both the legs and wings of a chicken! A lively discussion ensued about possible hidden factors that can mask student capabilities and understanding.

   b. A similar story comes from the city of Chicago. This time it's not language that blocks quality assessment, but the context and experience of the learners:

   On a statewide assessment, children were asked to identify the number of legs on a chicken. Children from rural communities had no trouble providing the *expected answer—two. City children, whose experience of chickens rested solely on those found in frozen packages in the grocery store, responded "four." When questioned, the children said that all the chickens they'd seen had four legs showing in the package.*

c. Another example is a test item about birds sitting on a fence. Once again, context and experience blocked quality assessment, and in fact, the answer expected by the test developers was incorrect when examining the realities of rural life:

Four birds were sitting on a fence. A farmer threw a stone that hit one of the birds. How many birds were left on the fence?

(Every farm child knew that if there was a stone thrown toward the fence, no matter how many birds were hit, all would fly away and none would remain.)

2. Ask participants to think of a time when they encountered a similar problem. Have participants examine the task (individually or in small groups) and discuss potential problems for students whose language and culture are not the same as that of the task developers. Use all or a few of the questions that follow (or create your own) to prompt individual reflection and then discussion:

- What do you think are some ways in which the task could pose problems for children of diverse cultures? Children whose first language is not English?
- As it is currently written, how much would you trust inferences about student achievement, knowledge, or skills from their responses to the task?
- How might this task be repaired to help gain a clearer picture of student ability? (**Extension**: Ask participants to underline parts of the task that could be changed to improve the task and then make changes that match your students' culture and environment.)
- What are some response options that do not require, or which minimize, written language requirements?

CHICKENS AND PIGS: LANGUAGE AND ASSESSMENT*

Problem: Ignacia and Cal went outside to feed the animals. They saw both chickens and pigs. Cal said: "I see 18 animals in all." Ignacia answered: "Yes, and all together they have 52 legs." How many chickens and how many pigs were in the yard?

* Adapted from NCTM

A5.3,H1
A5.3,O1

LABORATORY NETWORK PROGRAM

Activity 5.4

Equity Role Play

Purposes:

1. To explore personal beliefs, attitudes, and values by creating and playing the role of a student or teacher involved in potentially inequitable assessment activities

1. To understand that personal prejudices and biases can be very subtle and unintentional

2. To be aware that one's own issue with equity, access, and diversity is only a first step in developing and using assessments that are fair to all students

Uses:

1. In Chapter 2 to promote general awareness of assessment issues and alternative assessment situations

2. In Chapter 5 to create general awareness of the more subtle aspects of bias as we evaluate tasks and assessment situations for equity and accessibility

3. In Chapter 6 to review alternative reporting formats

4. In Chapter 7 to emphasize that professional development must address issues of access and equity over time

Rationale:

Issues of racism, sexism, classism, and any other "isms" involving bias, prejudice, discrimination, and other forms of inequity are emotional and uncomfortable topics for most people. Yet, ignoring problems, even unintentionally, only worsens situations. By playing "stereotypic" roles with some predefined characteristics, attitudes, and beliefs, participants can enter a safer environment in which to become aware of and reflect upon their own values and expectations, and beliefs about how children learn and how teachers and "outside" factors may influence and facilitate that learning.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens

- **Overheads A5.4,O1** to **A5.4,O5**
- Handouts of general character and specific role descriptions, assessment tasks, and observation and reflection forms (**Handouts A5.4,H1** to **A5.4,H13**)

Time Required:

90 minutes to 2¼ hours

Facilitator's Notes:

A. Logistics

For the opening and closing parts of the session, a room suitable for whole group presentation and discussion is needed. Because several plays will be acted out simultaneously, rather than one role play presented in front of an audience, break-out rooms may be needed. The room(s) should accommodate tables of four or five "cooperatively engaged students" with several "observing teachers" engaged with the groups.

Each play requires from five to eight participants, ideally involving six characters—two teachers and four students. Depending on the total number of participants, some plays may have one or two "Ringers" while others may need you, the facilitator, to play a student role.

B. Introduction (5 minutes)

Briefly introduce yourself and, if the group is not too large, have participants do likewise. Then provide a brief overview of the session goals, such as:

> *In this activity, you will explore your personal beliefs, attitudes, and values by creating and playing the role of a student or teacher involved in potentially inequitable assessment activities. Our goals are to understand that:*
>
> - *Personal prejudices and biases can be very subtle and unintentional*
> - *Awareness of our own issues with equity, access, and diversity is only a first step*

C. Opening (5 minutes)

You may wish to begin by relating a story from your own experience or knowledge of an incident involving discrimination, bias or prejudice, or introduce and then read a passage such as the paragraph below from Barbara Kingsolver's book *Pigs in Heaven* (Harper Collins Publishers, 1993). Show **Overhead A5.4,O1**—From *Pigs in Heaven.*

> *In the books* The Bean Trees *and* Pigs in Heaven, *a young woman, Taylor Greer, must overcome many difficulties in adopting a daughter, the Cherokee child named Turtle. Toward the end of* Pigs in Heaven, *the author writes: "Taylor is aware of being the white person here. Since her arrival in Oklahoma, she has felt her color as a kind of noticeable heat rising off her skin, something like a light bulb mistakenly left on and burning in a roomful of people who might disapprove. She wonders if Turtle has always felt her skin this way, in a world of lighter people."*

At this time, do ***not*** discuss the opening story or incident. This allows time for tension and personal reflection. In the upcoming "ice–breaker" activity (section F.2 below) individuals may share their own stories or reactions to the opener.

D. Setting the Stage (10-15 minutes)

1. Create a setting for participating in a role play by acknowledging people's concerns about sharing personal beliefs and the importance of building awareness not only of our own beliefs but also of those held by individuals who do not necessarily share our views or perspectives. To provide opportunity for further reflection (not sharing), you may wish to include information and rhetorical questions such as the following:

> *As educators, we live and work in a dynamic environment. As access to information grows easier, we become overwhelmed with the amount of information available and with the few resources we have to retrieve, understand, and act upon that information. We often wish to retreat into our classrooms, close the doors, and return to the "good old days" when everyone knew what was "right" and could confidently do the "right" things.*
>
> *Some of us may still believe that education, particularly in mathematics and science, can be "value-free" or we may hold onto the "one size fits all" school model, silently amazed that these "truths" are under attack as schools and districts restructure to meet the needs of their students and communities. We may find ourselves in the uncomfortable position of being "PI"—politically incorrect—in a society with national groups creating standards we personally may not accept, based on underlying assumptions contrary to our values or beliefs.*
>
> *To what extent does* ***everyone*** *believe that* ***all*** *children can learn? How is it that schools or districts embrace "inclusion" while allowing "tracked" classes to persist? How do we provide* ***all*** *children with access and opportunities to learn? How will teachers who believe intelligence is innate rather than dynamic provide "manipulative" activities and have children*

actively constructing their own knowledge and understanding? To what degree is it possible to treat all students equally **and** *equitably?*

With this role play-activity, we will work in small groups that are responsible for creating and maintaining safe environments for one another to express views and behaviors that are not necessarily "PC" (politically correct) and that are not necessarily our own (i.e., we each play a role).

2. Next, set the stage with the following scenario on which the role plays are based. Distribute **Handout A5.4,H1**—*Character Descriptions* and have participants review the major points of **"The Setting."**

   Next week, all grades in the Sunnyside Union Middle School will administer the district–mandated, highly touted "Concepts, Applications, and Communication Assessment" featuring individual performance activities as well as small group measures. Teachers and administrators are concerned that student performance will be considered when the district evaluates each school, and rumors abound that classroom–level data may be used to evaluate individual teachers' effectiveness.

   Veteran teacher Lee Crabtree and Kim Newcomer, a recent graduate of Teachers University Motivational School, have decided to focus on small group work with their math/science classes for the next week in preparation for the test. They will observe and critique one another's classes.

3. Tell participants that in today's session, each of them has the opportunity to "walk a mile in the shoes" of one of the Sunnyside Union Middle School (S.U.M.S.) characters that they will create based on certain information that will be provided.

E. The Players (10-15 minutes)

1. Explain that several plays will be performed simultaneously. Continue to refer to the handout to review the roles of **The Players:** Each play will have two teachers—Lee Crabtree and Kim Newcomer—who may display subtle yet racist, sexist, classist behaviors. Depending upon the number of participants, each play will involve from three to five middle-school students, working "cooperatively" on an assessment activity.

Amie	"Girl–next–door" type, sometimes seen as a "goody two-shoes"
Barnard	Highly confident, somewhat "elitist" young man
Carmen	"Happy–go–lucky" youth of mixed cultural heritage, often considered the "class clown"

L.J.	Recent immigrant and an "English Language Learner" (ELL)
"Ringer"	Highly individual student, often behaving in eccentric/erratic ways

2. Point out that several of the players are described with terms or labels that may be unfamiliar to some and/or interpreted differently by one another. In a few minutes, players will be grouped according to roles so they can work together as they "get into character." However, one new label—"English Language Learner"— should be presented to the group as a whole. Read or paraphrase the following text to illustrate the subtle messages we send when we use labels. Show **Overhead A5.4,O2**—*"ELL" or "LEP"?*

> *English Language Learners (ELLs) refers to students whose first language is not English, and encompasses both students who are just beginning to learn English (often referred to as limited English proficient or LEP) and those who have already developed considerable proficiency. The term underscores the fact that, in addition to meeting all the academic challenges that face their monolingual peers, these students are mastering another language—something too few monolingual English speakers are currently asked to do in U.S. schools. The term follows conventional educational usage in that it focuses on what students are accomplishing, rather than on any temporary "limitation" they face prior to having done so, just as we refer to advanced teacher candidates as "student teachers" rather than "limited teaching proficient individuals," and to college students who concentrate their studies in physics as "physics majors" rather than as "students with limited physics proficiency."**

When small groups are created, groups with six players work best (two teachers and the first four students described above). When necessary to increase the cast size, add a "Ringer." With fewer than six members in a play, omit the character Amie and have the character Carmen played as a female, who may take on some of Amie's characteristics.

F. Getting into Character (10-15 minutes)

1. Assign Parts

Determine the number of groups and sizes for each role play, and distribute the six or seven character handouts (Lee Crabtree, Kim Newcomer, Amie, Barnard, Carmen, L.J., and "Ringer" as needed) so that each participant has the sheets corresponding to the role to be created. It is not necessary that men play only male roles or that women play only female roles. Raising awareness involves trying "new" roles, seemingly distinct from one's own.

---

* (Quotation from footnote by LaCelle–Peterson and Rivera, "Is It Real for All Kids? A Framework for Equitable Assessment Policies for English Language Learners." *Harvard Educational Review,* vol. 64, no. 1, Spring 1994, p. 55)

Because it is important that *every participant play a role,* be sure that all participants, even the shy or reluctant ones, are assigned parts. When a role is played with little enthusiasm or even antagonism, participants can later debrief about their reactions when "forced" to do something or how they react when working with someone forced to "participate."

2. Breaking the Ice in Character Groups

Have participants meet with their character groups — i.e., others who will be assigned the *same* role. In these six or seven small groups, participants begin with an "ice breaker."

Have members share their reactions to the **A5.4,O1**—From *Pigs in Heaven* passage, or whatever passage or incident you used in the opening of this session. Or, you may show the **Overhead A5.4,O3**—*Ice Breaker* and have participants share with the others in their group an incident in which they experienced or witnessed discrimination, bias or prejudice (in the form of racism, sexism, classism, differences in language, or physical abilities) in an educational or work setting.

3. Discuss and Construct the Roles

Distribute character roles (**A5.4,H2** through **A5.4,H8**). (Within small groups, each participant receives the same role, and each small group receives a different role.) Participants should read through all the information provided on their character sheet. With their character groups, they should discuss the role and meaning of any terms. They may consider the body language as well as additional comments they might use during the plays to help portray their character.

Although several people may be playing the same role, individuals will build on their own experiences to "fill in the blanks" and construct their personal version of the character they will play.

4. Handouts

Be sure that the "teachers" in each group have the proper **Group/Student Observation Form**. Crabtree needs **Handout A5.4,H9**—*Group/Student Observations Form A*; Newcomer uses **Handout A5.4,H10**—*Group/Student Observations Form B*. Let Crabtree know if the students are to do the math or the science task. In part G, Crabtree will begin by distributing **Handout A5.4,H11**—*Group Assessment Task* for the "students" in the group.

(Depending on your participants, you may substitute the math or science task with another task that is suitable for active group work.)

Crabtree and Newcomer change observation roles after 10 minutes, and after another 10 minutes students complete **Handout A5.4,H12**—*Individual Self-Reflection Questions* as the teachers compare observation form notes.

G. Playing the Roles (25-30 minutes)

1. In each character group, have members count off (1, 2, 3, 4, etc.). Then have all the 1's go to one of the cooperative group tables (or break–out rooms), all the 2's proceed to another table/room, and so on until all the characters are with the appropriate cast in a role play.

2. Help groups keep track of time—20 minutes for student work on the task (10 minutes Crabtree observes and 10 minutes Newcomer observes) and 10 minutes for the student self-reflection while the teachers compare notes.

H. Group Debriefing and Self–Reflections (30-45 minutes)

1. Return to the large group setting for debriefing and discussions. Remind participants that each character may have received unfair treatment as well as behaved in a biased or prejudiced manner. Additionally, the assessment activity may raise issues of access for some of the students and fuel discussion about the teachers' abilities to assess equitably.

2. Show **Overhead A5.4,O4**—*Debriefing Questions.* Ask volunteers, one from each character role, to share their experiences, beginning with questions such as: *What specific "stereotypic" behaviors or characteristics did you incorporate in your role? Why? What responses did you expect to elicit from the players of other roles?* Other participants may describe how they played their roles as they interacted with the character being discussed.

3. Show **Overhead A5.4,O5**—*Self-Reflections.* Provide time for participants to write their personal reflections to questions such as: *Consider the role you played: What aspects of the character you created reflect your own personal beliefs or assumptions? How did attitude (yours or others) affect your ability to play the role effectively? What surprised you about this activity?*

4. Ask participants about any increased awareness that may have resulted from this activity. Have them talk about what steps can be taken to meet the challenges of providing meaningful and equitable educational experiences for all.

I. Extension

Distribute and discuss **Handout A5.4,H13**—*What Is Equity Anyway?* Invite participants to send their responses to Dr. Julian Weissglass, director of the Center for Educational Change in Mathematics and Science, University of California, Santa Barbara, CA 93106.

Character Descriptions

In this role play, you will "walk a mile in the shoes" of one of the S.U.M.S. (Sunnyside Union Middle School) characters that you create based on information about the situation and players.

The Setting

Next week, all grades in the Sunnyside Union Middle School will administer the district-mandated, highly touted "Concepts, Applications, and Communication Assessment" featuring individual performance activities as well as small group measures. Teachers and administrators are concerned that student performance will be considered when the district evaluates each school; and rumors abound that classroom-level data may be used to evaluate individual teachers' effectiveness.

Veteran teacher Lee Crabtree and Kim Newcomer, a recent graduate of Teachers University Motivational School, have decided to focus on small group work with their math/science classes for the next week in preparation for the test. They will observe and critique one another's classes.

The Players

Several plays may be performed simultaneously. Each play will have two teachers—Lee Crabtree and Kim Newcomer—who may display subtle yet racist, sexist, and classist behaviors. Depending upon the number of participants, each play will involve from three to five middle-school students, working "cooperatively" on an assessment activity. The student players include:

Amie	"Girl-next-door" type, sometimes seen as a "goody two-shoes"
Barnard	Highly confident, somewhat "elitist" young man
Carmen	"Happy-go-lucky" youth of mixed cultural heritage, often considered the "class clown"
L.J.	Recent immigrant and an "English Language Learner" (ELL)
"Ringer"	Highly individual student, often behaving in eccentric/erratic ways

The Plays

Each role play requires a minimum of five players: the two teachers—Crabtree and Newcomer—and three students—Barnard, Carmen (played as a female), and L.J. With six players, add the role of Amie. (Carmen may then be played as either male or female.) With seven or eight players, add one or two "Ringers".

Play begins with Crabtree administering a practice assessment to a small group of students and completing an observation form on the group. After a few minutes, Newcomer takes over the task administration and completes a second observation form. The plays end as students complete individual reflection forms and teachers compare observation notes.

Following the plays, you will debrief your experiences in a large group setting.

A5.4,H1

You are to play the role of **Lee Crabtree.** Use the following background and other information about Crabtree's beliefs and interactions to help create this character.

Background Lee Crabtree has 25 years of teaching experience, 20 at the high school level, and has been teaching math and science at S.U.M.S. for the last five years. Lee likes kids, a lot—and coaches after–school soccer without extra pay. Lee realizes that scores on district tests have been decreasing the last few years, and this year scores "count" toward some sort of school report card. Lee has just learned that the new "authentic" assessment the district has "cooked up" contains some cooperative group components. Since Lee's students sometimes work in groups, Lee is not too worried. They are pretty good kids, after all, and Lee is sure they can cooperate for a couple of hours. Lee, a monolingual English speaker, has watched the Sunnyside community change over the years, and is particularly frustrated to receive yet another new LEP student in class. The Sunnyside district has provided a few inservices on diversity, access, and equity, but Lee is tenured and doesn't feel the need to make any significant changes. Lee feels that non–native English speakers bring "excess baggage" into the classroom and represent just one more problem to overcome.

Interactions with Other Characters In general, Lee asks questions and expects instant answers, allowing very little wait time. Lee gives very little praise, typically an "uh-huh" or a nod to the girls, and often encourages boys to work harder. Lee is quite pleased when the boys do things right and often shows approval by giving them pats on the back. Lee fears that girls lack what it takes to succeed in science or math outside the classroom, and thinks that they will probably not need as much math and science in the future as boys will, except perhaps if they'd like to teach. Lee tends to compliment girls on their appearance and neatness, knowing that these are important aspects of an adolescent female's life.

Kim Newcomer	When Kim began teaching at S.U.M.S. this year, Lee took Kim "under the wing." Lee admires Kim's enthusiasm, but doesn't agree with all the learning strategies Kim uses.
Amie	Amie (Lee pronounces it "AY-mee") is the type of daughter Lee would like to have had. She is a quiet, skinny, cute little kid who always does her homework—perfectly and on time. Amie also helps out with classroom duties and even grades papers for Lee.
Barnard	"Barney," as Lee fondly calls him, is a "star" student and athlete. Lee is confident Barney will do anything he sets his mind to, and often pushes him to take the lead.
Carmen	Lee sometimes wonders how one child can always be so "happy–go–lucky" or smile so much. Usually Lee enjoys Carmen's "class clown" routine, and really doesn't expect much more from a child who missed a lot of school a few years ago and who seems to have some language problems, probably due to a non–English speaking family.
L.J.	Lee doesn't really know what native language it is that L.J. speaks, but does know L.J. must have some English–speaking skills because L.J. talks to other kids on the playground. Lee is unimpressed by any work L.J. does complete, and is impatient when L.J. doesn't understand directions or confuses classmates with incorrect explanations.
"Ringer"	Lee finds Ringer's individuality like a "breath of fresh air" and has become tolerant of Ringer's eccentric and sometimes erratic ways.

Role Decisions Begin the role play by distributing the "practice" **Assessment Task (A5.4, H11),** making sure students know if they're to do the mathematics or the science task. Complete the two–part **Group/Student Observation (Form A) (A5.4, H9)** as students work on the task. When necessary, "help" the students. After all, you want to be sure they'll do well on the upcoming district assessment. After 10 minutes, you are called to the front office and Kim Newcomer takes over the class. After another 10 minutes, you and Kim compare observation notes as students do their **Individual Self–Reflection Questions (A5.4, H12).** Before the play begins, decide:

- Will you play the role as a Mr., Mrs., or Ms. Crabtree? (Your gender and Crabtree's needn't be the same. Be sure the other players know and use your correct title.)
- What additional characteristics or dispositions will you portray in this role?
- What behaviors and mannerisms will you use?

A5.4,H2

You are to play the role of **Kim Newcomer.** Use the following background and other information about Newcomer's beliefs and interactions to help create this character.

Background Kim Newcomer is 24 years old and has just completed both an undergraduate and graduate degree at the prestigious Teachers University Motivational School. Newcomer anticipated this first real teaching job with great enthusiasm, and now is experiencing some doubt about a future in teaching. Kim has just learned that district scores on important tests have been on the down side the last few years. If Kim's classes don't do well, rumors are that Kim will not be asked to return to S.U.M.S. next year. This district test is catching Kim off-guard. Lee had said there was nothing to worry about, but now Kim discovers that other teachers have been having their classes practice for tests like this the entire year. Kim's students often work collaboratively to solve group problems, but Kim is concerned the students won't do well on the writing parts where they must tell about their thinking—so many of them have better ways of demonstrating their knowledge. Kim is thankful that Lee agreed, albeit reluctantly, to practice some non-standard tasks that require students not only to work cooperatively but also to write. Kim knows that the teacher observations they will do will help both teachers and students see areas for improvement.

Interactions with Other Characters At the University, Kim learned about "learning styles" and has tried to apply that knowledge in instructional situations. With "linguistic" learners, who like to read, write and tell stories, Kim often has them talking and explaining their thinking. For "visual/spatial" learners, Kim suggests drawing and using manipulatives to solve problems. When getting small groups to work cooperatively and collaboratively, Kim is unsure what to do with students who prefer working alone as well as with those who monopolize the group.

Lee Crabtree	Kim has been grateful to have had Lee to show her the ropes around S.U.M.S. However, Kim thinks Lee should leave the textbook and use more real-world problems. Students complain they are bored in Lee's classes and that Crabtree is always "on their case."
Amie	Kim uses Amie's given name, Amica (pronounced AH-mih-cuh), and fears that she is beginning to run with a "fast" crowd. Kim thinks Amica often hides her limitations by being helpful rather than asking for help. Kim feels a "distance" between student and teacher.
Barnard	Kim cannot keep far enough away from Barnard and finds him pompous and arrogant—using good looks and "charm" to get others to do the work.
Carmen	Kim knows that Carmen had leukemia in third grade and missed a lot of school during two years of chemotherapy. Kim sees Carmen's smile as a mask to hide true feelings. Realizing that bilingual kids cannot always receive necessary speech therapy, Kim was able to get Carmen into an LD program for the "learning disabled."
L.J.	Kim doesn't know L.J. too well, but thinks this student might be a "kinesthetic" learner because L.J. seems active in playground sports. Kim sees L.J.'s bullying as a defense to being a new kid.
"Ringer"	Ringer uses and confuses Kim. It's difficult to know what to expect from this active student. Sometimes Ringer's work is excellent, sometimes Ringer just waits to be told what to do.

Role Decisions In the role play, you quietly observe as Lee begins the administration of the practice assessment. After 10 minutes, tell Lee there is a call at the front office. You then complete a separate two-part **Group/Student Observation (Form B) (A5.4, H10)** as students continue to work on the task. Feel free to help the students, since this is for practice. After 10 minutes, collect the students' responses and give each student an **Individual Self-Reflection Questions (A5.4, H12)** sheet to complete. During this time, you and Lee compare the ratings you gave students on the two observation forms. Before the play begins, decide:

- Will you play the role as a Mr., Mrs., or Ms. Newcomer? (Your gender and Newcomer's needn't be the same. Be sure the other players know and use your correct title.)
- What additional characteristics or dispositions will you portray in this role?
- What behaviors and mannerisms will you use?

A5.4,H3

You are to play the role of **Amie.** Use the following background and other information about Amie's beliefs and interactions to help create this character.

Background Amie—pronounced AH-mee, and "short" for Amica (pronounced AH-mih-cuh)—comes from a "mixed heritage" home. Her mother is Portuguese and her father Russian. However, her parents have insisted Amica speak only English at home and school, and they long ago stopped using their native languages, trying to become Americanized as quickly as possible. Although she has been called AY-mee and Uh-MEE since she started school, Amie now wants people to use her given name, Amica, but she is afraid to tell her teachers and some of her classmates for fear they won't like this unusual name and will think she is weird.

Amie works very hard in school, but learning is difficult for her. She is meek, and seems unable to get the teacher's attention even when she knows what questions to ask. Her new friends tell her that it's better not to be too smart; boys don't like smart girls anyway. However, because education is so very important to Amie's parents (they punish her when her grades are below B), Amie tries to keep her grades up.

Interactions With Other Characters Amie is a scared, nervous, skinny adolescent, small compared to her classmates. She has begun spending more time on her appearance so that boys will like her. Amie is still concerned about her grades and wants the teachers to like her, too. So she continues to volunteer to help them out with grading and clean-up jobs. She also enjoys group work in class and often volunteers to be "recorder." This way, she gets to see how other kids do their work and that helps her learn.

Lee Crabtree	Amie feels that Lee Crabtree is a bit prejudiced toward people who are different. She wants to keep her heritage private. Amie volunteers to check homework and do other jobs for Crabtree in order to please her teacher and her parents.
Kim Newcomer	Amie thinks Kim Newcomer is a bit "stand-off-ish" and probably doesn't approve of Amie's new group of popular friends. It seems nothing Amie does really meets with Newcomer's standards.
Barnard	Amie is crazy about Barnard and gets angry when kids call him "Brainyard." She thinks he is not only the cutest, but the smartest guy at S.U.M.S. Amie is often at a loss for words around Barnard.
Carmen	Amie envies the happiness she thinks Carmen smile represents. She knows this classmate is not too bright, so she helps out and tries to make lessons easier whenever she can.
L.J.	Amie empathizes with L.J.'s struggle to fit into a school world very different from home life. But, Amie takes it personally that L.J. doesn't seem to like her very much, and tells L.J. to "just shut up" whenever L.J. complains or says mean things to other students.
"Ringer"	Amie wishes she could be more like Ringer and tries to be Ringer's friend. Sometimes Ringer seems so smart. Her new friends like Ringer because of the wild and dangerous things Ringer does.

Role Decisions In the role play, Amie puts on a "goody two-shoes" behavior with the males but is competitive with the females. Reluctant to begin the assessment task on her own, Amie is unsure of the meaning of all the words and afraid to show she doesn't understand. She begins by asking everyone for the correct spellings of their names and volunteers to take notes for the group.

Before the play begins, decide:

- What additional characteristics or dispositions will you portray in this role?
- What behaviors and mannerisms will you use?

A5.4,H4

You are to play the role of **Barnard.** Use the following background and other information about Barnard's beliefs and interactions to help create this character.

Background Barnard is the product of high–achieving parents who started off without much money, but managed to send him to private school and summer camps until fifth grade, when they moved to the Sunnyside district. Barnard's father works six days a week and seldom sees the family. Barnard directs his disdain for women, particularly his mother, toward most females, teachers as well as classmates. Barnard loves physical activities, especially tennis and swimming, and seems driven toward his goals to win at sports and be the best at school. He prides himself on his appearance and intelligence, and has aspirations of attending a prestigious university.

Interactions With Other Characters Where school is concerned, Barnard prefers to work alone. He figures it's easier to do it right by yourself than to wait around until others "get it." Sometimes they're just so slow! When forced to work with a group, Barnard is patronizing and typically plans to re–do the final product by taking it home to add a few "touch-ups." Inwardly Barnard is pleased that others seem to want him to be leader, but becomes annoyed when he thinks they expect too much from him. He hates being called Barney, like some cartoon character, and smugly dismisses those who call him "Brainyard," with "What's the matter? Jealous?"

Lee Crabtree	Barnard likes to keep Crabtree happy, and generally enjoys Crabtree's classes since most work is done individually.
Kim Newcomer	Newcomer's "innovative" lessons can be fun, but all these "reflections" seem like a waste of time to Barnard; and, of course, Newcomer is always putting them in groups. But, Barnard thinks Newcomer is rather easily manipulated, so it's not a problem.
Amie	Barnard calls her AH-mee (rhymes with "mommy") and otherwise barely notices this little mouse. She's just about useless in groups and never contributes much, except she does take good notes.
Carmen	Carmen seems not to be "playing with a full deck." Whether it's for a game or a group project, this is not someone Barnard would choose for a teammate. This kid's a loser.
L.J.	Barnard thinks L.J.'s family should go back to wherever it is they came from. Kids like this just slow the group down. It surprises Barnard when L.J. bullies other kids or acts like a "know–it–all."
"Ringer"	Sometimes this kid has great ideas, and other times Ringer is just "off the wall." Basically, Barnard tolerates Ringer, until the name–calling or "finger drums" begin.

Role Decisions In the role play, Barnard wants everyone to just try the assessment task themselves and work alone. He complains, "Where's the computer when you need it?" He wants to use a spreadsheet program (like his dad does) to work with the numbers and "crunch this out in 5 minutes." If Barnard were doing this at home, he brags that he would even do a simulation program with graphics.

Before the play begins, decide:

- What additional characteristics or dispositions will you portray in this role?
- What behaviors and mannerisms will you use?

A5.4,H5

You are to play the role of **Carmen.** Use the following background and other information about Carmen's beliefs and interactions to help create this character.

Background Carmen is a bilingual child whose extended family speaks only Spanish in the home. People are often enchanted by this child who has a warm sense of humor and understands kidding in a fun way. At age eight, Carmen was diagnosed with leukemia and received chemotherapy for the next two years, missing a lot of school. Though the leukemia is in remission, Carmen has blood tests twice a year, and still has memory gaps, probably the result of the chemotherapy. Before the leukemia, Carmen did average work in school and was in a bilingual class in first and second grades. Now, language has become an issue in school in more ways than one. Carmen cannot receive as much speech therapy as may be warranted due to the effects of chemotherapy. As an "LD resource" student, Carmen receives language development in comprehension and memory, but not speech correction. With the speech therapist, Carmen is also learning to work with facial expressions. It seems Carmen always wears a smile, whether having a problem or having fun. Carmen hadn't learned to change facial expressions to reveal inner feelings. Also, Carmen is learning strategies for actions most people take for granted, such as going up to a teacher and asking for help.

Interactions with Other Characters As class clown, Carmen receives a lot of attention. Others seem to enjoy the off-handed, sometimes sarcastic remarks, even when they're delivered in Spanish.

Lee Crabtree	Carmen would like to please "The Crab" and often plays around in Crabtree's class. Carmen has not learned to let Crabtree know when the work is too difficult or the directions confusing.
Kim Newcomer	With Newcomer, Carmen tries to "lose the smiley face" because Newcomer seems to understand about Carmen's language and facial difficulties. However, Carmen continues to play "happy-go-lucky," not ready to ask for help when tasks seem too difficult.
Amie	Carmen considers Amie a bit of a pain. She is always trying to be overly helpful and protective. Sometimes Carmen just can't stop making sarcastic remarks in front of her.
Barnard	Carmen enjoys fooling around with "Brainyard" to the point of being too annoying. When Carmen does try to settle down to business, behavior toward "Barney" can somehow be just as annoying.
L.J.	With L.J., Carmen seems more relaxed. Perhaps they work well together because there's not much verbal interchange, mostly just working with materials. Unfortunately, they seldom get correct answers or do problems correctly, sometimes even confusing others.
"Ringer"	Carmen gravitates toward Ringer, who tends to bring out the clown in this classmate. Together, they can get into trouble, particularly when they target Barnard for teasing.

Role Decisions In the role play, Carmen begins clowning around, but becomes interested in the assessment task and wants to do really well to impress the teacher. Before the play begins, decide:

- Decide whether you will play the role as female or male. Your gender and Carmen's needn't be the same. However, if there are fewer than six players for your group's role play, the character of Amie will be omitted. You should play Carmen as a female so that there are enough females in the cast. Be sure the other players know how you are playing the role so they can play their roles accordingly.
- What additional characteristics or dispositions will you portray in this role?
- What behaviors and mannerisms will you use?

A5.4,H6

You are to play the role of **L.J.** Use the following background and other information about L.J.'s beliefs and interactions to help create this character.

Background Sunnyside is L.J.'s second American school this year. Previously, L.J. was an excellent student and was just learning English, but in a war-torn country. L.J. longs for his old friends and real home, but worries the family might have to go back before the fighting is over. L.J.'s parents and a few neighbors from their homeland often discuss that possibility when chatting in their native language. L.J. has low respect for women, and can't say one nice thing about them. With a pregnant mother and three younger sisters, L.J. does a lot of the work around the home, often shopping or doing errands the mother is reluctant to do. As the oldest child, L.J. is idolized by these younger sisters and would do anything for them, but L.J. also envies their ability to make friends and fit into this U.S. world.

Interactions with Other Characters L.J. speaks enough English to get by, but is embarrassed when doing any oral reading. L.J. tends to bully others and will make up stories to come across as much brighter than the other kids think. On the playground, L.J. can be somewhat aggressive, but usually wants to play with the others. Preferring to be actively engaged in small group work when materials are involved, L.J. sometimes "hogs" the supplies trying to show others what they should be doing.

Lee Crabtree	L.J. thinks Crabtree deserves the nickname "The Crab" and tries to avoid the teacher as much as possible, often saying "I got it" and refusing to admit that directions are unclear when Crabtree asks.
Kim Newcomer	L.J. senses Newcomer's anxiety over teaching, particularly where these upcoming tests are concerned, and wants to please this teacher.
Amie	L.J.'s aggressive behavior definitely extends to AY-Mee, seeing this little, skinny weakling as just another female who can't take care of herself. Amie seems to bring out the worst in L.J.
Barnard	L.J. is particularly jealous of Barnard. If the tables were turned, and "Brainyard" were the new kid in L.J.'s homeland, "Barnyard" would know how it feels to be a smart kid without ways to communicate.
Carmen	L.J. feels a kinship with Carmen and enjoys when they can work together. However, when Carmen wants to be with others as well, L.J. gets a bit possessive, wanting exclusive attention from Carmen.
"Ringer"	L.J. tends to ignore Ringer, considering the classmate to be a bit of a show-off without too much in the "smarts" department. However, it can be a different story when Ringer gets involved with materials to work on problems. Then L.J. thinks they can be best friends.

Role Decisions In the role play, L.J. looks over the task and then around at the classmates, displaying visible irritation with the situation. With comments such as "This stupid! Why do it?" L.J. seeks to avoid working on the problem, preferring to just go out and play. After a few minutes L.J. quietly admits not understanding what some of these words mean. L.J. then insists that if they have to do this problem they should at least have some materials and supplies to really try it out.

Before the play begins, decide:

- Will you play the role of L.J. as a female or male? (Your gender and L.J.'s needn't be the same.) Decide how you will let the other players know whether L.J. is a boy or girl, so they may play their roles accordingly.
- What additional characteristics or dispositions will you portray in this role?
- What behaviors and mannerisms will you use?

A5.4,H7

You are to play the role of **"Ringer."** Use the following background and other information about Ringer's beliefs and interactions to help create this character.

Background "Ringer" chose this nickname early in school, learning that a "ringer" can be something that "throws people off" and Ringer does that well. Overly active to the point of occasionally being out of control, Ringer has been on medications since age nine. On reports, teachers have used phrases like "off in own little world," "keeps people on the outside," "gets all upset over the least little things," "does anything wants to, at any time." Ringer enjoys reading maps and charts, and when activities call for drawing, building, designing, or creating, Ringer does well. With a vivid imagination, Ringer often pictures things in vivid colors and sometimes just feels music inside, which tends to annoy others when Ringer starts drumming on the desk. Problems begin when Ringer just needs to get up and move. Ringer gets bored easily in class—it's hard to sit still very long.

Interactions with Other Characters Ringer works just as well alone or with others. Wanting to be liked by everyone, Ringer will sometimes do outrageous things to get attention.

Lee Crabtree	Ringer feels that Crabtree is too up–tight and enjoys giving the teacher a hard time, but in a playful way. Sometimes Crabtree will let Ringer get away with anything.
Kim Newcomer	Newcomer's another pushover. Trying to "find the best in everyone" Newcomer will sometimes give away answers if you just ask enough times or wait long enough before you get to work.
Amie	Ringer knows that AH-mee, or Amica (pronounced AH-mih-cuh) as her new friends call her, likes it when Ringer shows off. Amie provides a great audience and is easily impressed even when Ringer says or does stupid things.
Barnard	Ringer likes to make Barnard angry by calling him "Barnyard" or "Brainyard," and laughs when Crabtree calls him Barney, like that character in kids' cartoons. Ringer is a bit jealous and resents it whenever Barnard takes the leadership role.
Carmen	Ringer can't quite figure out "Smiley" but tends to look out for the classmate if others are mean or start to tease the kid. Sometimes they can get into all sorts of trouble together.
L.J.	Ringer doesn't have much to do with this classmate unless there's work to be done that requires more activity than language. L.J. is too much of a bully for Ringer's tastes, and is better left out than welcomed in when it comes to group activities.

Role Decisions In the role play, there may be more than one "Ringer" if group size requires more than five students. In that case, one of the ringers will need to change nicknames. During the assessment, Ringer gets involved right away with the task and insists upon looking for materials and supplies to build models. Ringer is quite vocal and would like to take the lead for the group.

Before the play begins, decide:

- Will you play the role of "Ringer" as a female or male? (Your gender and Ringer's needn't be the same.) Decide how you will let the other players know whether Ringer is a boy or girl, so they may play their roles accordingly.
- What additional characteristics or dispositions will you portray in this role?
- What behaviors and mannerisms will you use?

A5.4,H8

Group/Student Observations Form A (for CRABTREE)

Rate the group as a whole on the upper part of this form and rate individual work on the lower part.

Group ____________________ **Date** ____________________

Activity ____________________

	frequently	sometimes	never
Individual Ideas: Within the group, do members...			
- **Clarify their own ideas?**	☐	☐	☐
- **Evaluate one another's ideas?**	☐	☐	☐
- **Compare alternatives?**	☐	☐	☐
Communication: Do members in the group...			
- **Fill the role of both "talker" and "listener"?**	☐	☐	☐
- **Synthesize and summarize the group's thinking?**	☐	☐	☐
- **Express ideas clearly and effectively?**	☐	☐	☐
Cooperation: Does the group...			
- **Divide the task(s) among the members?**	☐	☐	☐
- **Agree on a plan or structure for tackling the task?**	☐	☐	☐
- **Take time to ensure that all understand the task?**	☐	☐	☐
- **Provide support for each member?**	☐	☐	☐
- **Allow for development of leadership?**	☐	☐	☐

Rate each student on CONCEPT UNDERSTANDING (indicate which students are at each of the three levels of development)

NOT UNDERSTANDING	DEVELOPING	UNDERSTANDING AND APPLYING
a Does not routinely model the concept correctly **b** Cannot explain the concept **c** Does not attempt problems **d** Does not make connections	**e** Demonstrates partial or satisfactory understanding **f** Can demonstrate and explain using a variety of modes (e.g., oral, written, objects, model, drawings, diagrams) **g** Is starting to make how-and-why connections **h** Relates concept to prior knowledge and experiences **i** Can create related problems **j** Accomplishes tasks, though with minor flaws	**k** Correctly applies rules or algorithm on how to manipulate symbols **l** Connects both how and why **m** Can apply the concept in new or problem situations **n** Can see and explain connections **o** Accomplishes tasks and goes beyond
Students:	**Students:**	**Students:**

Adapted from Beyer, Ann. "Assessing Students' Performance Using Observations, Reflections, and Other Methods." In *Assessment in the Mathematics Classroom,* 1993 Yearbook of the National Council of Teachers of Mathematics, edited by Norman L. Webb and Arthur F. Coxford, pp. 111-120. Reston, VA: The Council, 1993.

A5.4,H9

Group/Student Observations Form B (for NEWCOMER)

Rate the group as a whole on the upper part of this form and individual work on the lower part.

Group ______________________ **Date** ______________________

Activity ______________________

	frequently	sometimes	never
Individual Ideas: Within the group, do members ...			
• **Clarify their own ideas?**	☐	☐	☐
• **Evaluate one another's ideas?**	☐	☐	☐
• **Compare alternatives?**	☐	☐	☐
Communication: Do members in the group ...			
• **Fill the role of both "talker" and "listener"?**	☐	☐	☐
• **Synthesize and summarize the group's thinking?**	☐	☐	☐
• **Express ideas clearly and effectively?**	☐	☐	☐
Cooperation: Does the group ...			
• **Divide the task(s) among the members?**	☐	☐	☐
• **Agree on a plan or structure for tackling the task?**	☐	☐	☐
• **Take time to ensure that all understand the task?**	☐	☐	☐
• **Provide support for each member?**	☐	☐	☐
• **Allow for development of leadership?**	☐	☐	☐

Rate each student on SUMMARIZING AND INTERPRETING RESULTS (indicate which students are at each of the three levels of development)

NOT UNDERSTANDING	DEVELOPING	UNDERSTANDING AND APPLYING
a Makes no attempt to summarize or describe data **b** May answer simple questions related to data if prompted **c** Cannot communicate results in rudimentary form	**d** Summarizes and describes data appropriately **e** Can generate and answer questions related to data **f** Can communicate results in rudimentary form	**g** Draws valid conclusions and interpretations **h** Makes generalizations **i** Communicates results clearly and logically
Students:	**Students:**	**Students:**

Adapted from Beyer, Ann. "Assessing Students' Performance Using Observations, Reflections, and Other Methods." In *Assessment in the Mathematics Classroom,* 1993 Yearbook of the National Council of Teachers of Mathematics, edited by Norman L. Webb and Arthur F. Coxford, pp. 111-120. Reston, VA: The Council, 1993.

A5.4,H10

Group Assessment Task

Your Name ______________________________ **Date** ____________________

Group ______________________________

Work with your group to find a solution to the math or science problem below. (Ask the teacher if you are not sure which task your group should complete.) Use your own words to write about your solution. Include drawings, tables, diagrams, and so on. Be sure to tell how you solved the problem and how you know your solution is a good one.

Math Task:

Which is the better fit—
A round peg in a square hole?
Or, a square peg in a round hole?

Science Task:

How does changing the weight on the end of a pendulum affect the period of the pendulum? Make a prediction. Then design an experiment you would do to test your prediction.

Individual Self-Reflection Questions

Student ______________________ **Date** ______________

Activity ______________________

Use the following questions to help you look back and describe your thinking as you worked toward a solution to the problem.

1. What did you do when you first saw the problem? What were your thoughts?

2. What strategies did you use to solve the problem? How did they work out?

3. Suppose you tried an approach that didn't work and you had to stop and try another approach. How would (did) you feel?

4. In what ways did you check you solution? How sure did you feel that it was reasonable?

5. What were your feelings about working with your group?

6. What new mathematics or science have you learned or what have you understood better about mathematics/science as a result of working this problem?

The following selections are from Julian Weissglass in the newsletter *Many Waters* (Fall 1993), p. 11-12.

What is equity anyway?

A complete definition of educational equity is elusive because the subject is extremely complex. Since most people have a sense (from their personal experience) of the effects of prejudice and discrimination on peoples' lives it is not necessary to provide a formal definition in order to address it effectively. A 'working definition' that I use is:

Educational equity is an ongoing process of increasing our own and society's capacity and commitment to:

1. Completely respect individuals as complex thinking and feeling humans of different genders and with different sociocultural and class backgrounds and values

2. Provide the necessary resources to assist people to learn, but also to overcome the effects on their ability to learn of any mistreatment—whether the source be accident, family, or society

What is your working definition of equity? Write and we will start a dialogue in this newsletter.

Perspectives on Equity

1. No one is born prejudiced. All forms of bias, from extreme bigotry to unaware cultural biases, are acquired (actually imposed on the young person) and are dysfunctional.
2. All humans are very much alike. We are one species.
3. In many societies the assumptions, values, and practices of people and institutions from the dominant culture serve to the disadvantage of students from the non-dominant culture.
4. Racism is more than the sum of individual prejudices. It is a complex social and economic phenomena that affects all aspects of U.S. society—in particular, education.
5. Racism, sexism, and classism are serious issues facing U.S. society and education that are usually not discussed. Talking about them is necessary, not to lay blame, but to figure out better ways of educating our children.
6. In the U.S., the lack of acceptance and support of a variety of forms of leadership is an impediment to the development of teacher leadership among people of color, women, and the working class.
7. It will be necessary to improve alliances between white educators and educators of color, between males and females, and between people of different class backgrounds, in order to make progress on this very complex problem.
8. Discussing and gaining new understandings about the existence and effects of bias and discrimination will usually be accompanied by strong emotions.
9. Changed attitudes and actions will be facilitated if we are listened to attentively and allowed to release our emotions as we attempt to make sense of what we and others have experienced (are experiencing).

What are your thoughts about this?

You may send your responses to Dr. Julian Weissglass, Director of the Center for Educational Change in Mathematics and Science, University of California, Santa Barbara, CA 93106.

A5.4,H13

From Pigs in Heaven...

"Taylor is aware of being the white person here. Since her arrival in Oklahoma, she has felt her color as a kind of noticeable heat rising off her skin, something like a light bulb mistakenly left on and burning in a roomful of people who might disapprove. She wonders if Turtle has always felt her skin this way, in a world of lighter people."

Barbara Kingsolver, *Pigs in Heaven*
Harper Collins Publishers, New York, 1993

A5.4,01

LABORATORY
NETWORK PROGRAM

"ELL" or "LEP"?

"English Language Learners" (ELLs)

- Refers to students whose first language is not English
- Includes students who are just beginning to learn English ("limited English proficient" or "LEP") as well as students who have already developed considerable proficiency
- Focuses on what students are accomplishing, rather than on any temporary "limitation" they face prior to having done so

Advanced teacher candidates—"student teachers" or "limited teaching proficient individuals"?

College students who concentrate their studies in physics—"physics majors" or "students with limited physics proficiency"?

A5.4,O2

LABORATORY NETWORK PROGRAM

Ice Breaker

Have you ever...

- **Experienced discrimination, prejudice or bias firsthand, in an educational or work setting?**

- **Witnessed behavior that is racist, sexist, or classist?**

Share with the others in your group an incident in which you experienced or witnessed discrimination, bias, or prejudice in an educational or work setting.

A5.4,O3

LABORATORY NETWORK PROGRAM

Debriefing Questions

What specific "stereotypic" behaviors or characteristics did you incorporate in your role?

Why?

What responses did you expect to elicit from the players of other roles?

A5.1,O4

LABORATORY
NETWORK PROGRAM

Self-Reflections

What aspects of the character you created reflect your own personal beliefs or assumptions?

How did attitude (yours/others) affect your ability to play the role effectively?

What surprised you about this activity?

A5.4,O5

LABORATORY
NETWORK PROGRAM

Activity 5.5

Questions About Culture and Assessment

Purposes:

1. To expand awareness of cultural factors that affect students' ability to clearly communicate their learning during alternative assessments and our ability to interpret the quality of their work
2. To promote serious reflection and discussion of questions that are sometimes given little attention in assessment planning
3. To bring up additional questions and unique answers about culture and assessment

Rationale:

When alternative assessments are planned and implemented as part of systemic initiatives, there are profoundly important questions to ask and answer about culture. Examining the interaction of culture and assessment is essential in order to assure that children have the best chance possible to create a rich portrait of their knowledge and capabilities.

Uses:

This activity can be used alone or with other Chapter 5 activities to focus on equity issues and assure that they are carefully and thoughtfully considered. Related activities include:

- **Activity 5.3:** *Chickens and Pigs: Language and Assessment*
- **Activity 5.4:** *Equity Role Play*

Materials:

- Overhead projector, screen, blank transparencies, and transparency pens or chart paper and pens
- **Handout A5.5,H1**—*Critical Questions*

Time Required:

90 minutes

Facilitator's Notes:

Answers to the questions outlined in **Handout A5.5,H1** may differ where participants and students share a common culture. The activity can then become primarily a reminder of cultural factors to be consciously considered in alternative assessments. When participants bring separate cultural perspectives and experiences from those of their students, answers to critical questions need to be crafted in partnership with students and others whose knowledge of the culture(s) can be tapped. The language(s) of the discussions in this environment should be based on the needs of participants to communicate in the language in which they can best express their thoughts.

1. One way to begin this activity is to open the session with one of the first questions—**What's the vision of the community for the future of its children?**

   In situations where parents and community members are invited to take part, this is a starting point that builds bridges between educators and families. Often the dreams that teachers, administrators, families, and communities have for their children are strikingly similar. Small role groups can brainstorm and share their answers, with the facilitator helping to emphasize the common dreams.

2. Following this consensus building, participants can be given handout **A5.5,H1**—*Critical Questions* and asked to select one or several questions for reflection and discussion. If your purpose is to create awareness of the multiple ways in which people answer these questions, cross-cultural groups might be effective. In all instances, individuals should be given the opportunity to self-select both the grouping preferred and the question(s). When participants work in "same culture" groups, you can use a jigsaw strategy that draws on the knowledge of each group and makes members responsible for sharing answers across cultures following the "same culture" discussions.

   Groups can consider all questions or divide up the questions across groups depending on time available and whether or not this is an ongoing process for improvement.

3. When answers to critical questions are to be shared more widely and formally beyond the group, there are additional considerations. Some participants may be free to speak for their culture, others may not. In several groups engaged in this process, members wished to consult with cultural leaders and respected elders to validate their preliminary responses.

Issues like the degree of elaboration that is considered sufficient and appropriate often arise in the midst of this process. For example, some groups want a lot of detail:

The Role of Context:

High context cultures tend to use communication strategies in which most of the essential meaning is embedded in physical and non-verbal means. **Low context** cultures, by contrast, tend to use communications in which meaning is made explicit through detailed and sometimes elaborate verbal or written messages. What are the implications for assessment? Members of cultures in which communication tends to be high context interpret "over-elaboration" as an indication that the speaker believes they don't "get it." When the home culture values high context communication, the pressure for elaboration within school assessments presents a significant challenge to students. How do we build bridges for children across contexts so that they can be successful within the school culture without dishonoring their home culture?

When Does Positive=Negative?

Feedback about the quality of student work is greatly affected by the role of feedback in a culture. Positive feedback provided publicly to individuals or groups is inappropriate in some cultures. Overly positive feedback in some cultures is used to call attention to the weaknesses of the work. Awareness of the role of feedback, appropriate ways to give it, who may give it, and the power of positive or negative feedback is essential.

What's the Role of Questioning?

Who asks questions differs widely across cultures. Are student questions of the teacher perceived as challenges to authority? Can distinctions be made between questioning to learn and questioning to challenge? How do efforts to promote critical thinking impact culture and family relationships?

4. Facilitators and participants are encouraged to brainstorm additional critical questions. Some of the most fruitful discussions come from the questions raised by participants. This is an activity which can be very hard to close. Depending on your time limits, you can close by emphasizing that the handout is food for further thought and action.

Critical Questions

This is a preliminary and incomplete list of questions about the complex relationship between assessment and culture. Many overlap. Some may be critical in some cultures, and relatively minor in others. Your additions, reactions, and suggestions are very welcome. What are your culture-assessment issues?

CULTURE AND ASSESSMENT

Assessment Question	Related Questions and Issues
What's worth knowing? Doing? Caring about?	What knowledge is essential across cultures? What knowledge is essential within a culture? To what extent should/must "universal" knowledge take precedence? What's the vision of the community for the future of its children? How do we determine what's worth knowing in a culture? Who determines? What are the essential skills and habits of mind **across** cultures? What knowledge, skills, cultural practices, and values from within the students' culture must be assessed? Whose role is it to provide knowledge from within the culture? To what extent can schools take part without usurping authority in the culture? How do we develop quality assessments for deep cultural knowledge? What language or languages should be used? Who must be involved? To what extent is knowledge "owned" in the culture? What implications does this have for making the knowledge overt through instruction and assessment?

A5.5,H1

Assessment Question	Related Questions and Issues
What are the purposes for assessment?	What do teachers want assessments to do for them? For their students? For parents?
	Who are the key receivers of assessment information in the culture?
What are key qualities of fine work?	How are criteria determined? Who determines the criteria used for assessing student work?
	What does quality look like in the culture?
	Where complex cultural practices and knowledge are involved, how can teachers from another culture determine quality? Must assessment be carried out only from within?
How can students be prepared to display vivid images of their learning?	What are powerful metaphors and parables of learning and assessment in the culture?
	Are visual images useful? Is the culture one that values oral expression?
	What are the communication patterns and preferences? To what extent should assessment tasks honor those patterns? To what extent and at what level(s) should assessment tasks require students to stretch themselves beyond their comfort?
How will we know they're learning?	What are appropriate ways for children to display their learning in the culture?
	What language will assure students opportunities to truly display their learning? What assessment purposes require the use of students' home language? A second language?
	To what extent should assessments require students to be multicultural?
	Is self-reflection a universal habit of mind? Is the revealing of self-reflection to others appropriate? Or is it more related to personality?

A5.5,H1

Assessment Question	**Related Questions and Issues**
	What are the key differences and similarities between the culture of the home and community, and the school culture?
What's a quality assessment task?	Should tasks be "culture free"? How much context is essential? Should we begin with local context and then design tasks that require knowledge of other cultures?
	What kinds of tasks are common in the culture?
	Is the culture one that emphasizes high-context communication? What are implications for assessment? For gauging the quality of student work?
	Who can work together in a group? Who cannot be in the same group?
	Will grouping by gender help or hinder?
Do you see what I see?	Are perceptions of quality the same across cultures?
	What are indicators of quality in the culture?
How should students be given feedback about the quality of their work?	What are appropriate forms of feedback in the culture? Is individual, public feedback viewed as positive? Very negative? Is non-public individual feedback effective? Acceptable?
	Is group feedback preferred? Are there strategies for providing individual feedback for work done in groups?
	Is verbal feedback effective? Written feedback?
	Who can give feedback? Who cannot? Is peer assessment acceptable?
	Is feedback in the culture focused on the strengths of the work? Is it focused on weaknesses? Do children from a culture that provides negative feedback respond to positive feedback?

Chapter 6

Using Alternative Assessment in Grading and Reporting

What's in this Chapter?

While our concepts about learning and teaching have grown and broadened in recent years, grading and reporting systems often continue to focus on the quantity of knowledge learned. Grades are still used to identify which students know most and which know least. Report cards based on this view of grading use a single number or letter to categorize students from most to least knowledgeable with cutoff scores that determine A's, B's, etc. So far, the chapters in this *Toolkit* have examined issues surrounding quality mathematics and science alternative assessment—clear and significant learning targets, meaty tasks, well-developed performance criteria, equity and bias issues, and more. While all of these are essential for quality assessment, using the information they provide about student learning to grade and report has its own serious considerations. This chapter addresses grading and reporting issues and pose questions that need to be carefully considered as we integrate alternative assessments into the overall educational system.

Chapter Goals

1. Heighten awareness of the importance of aligning grading and reporting strategies with valued learning outcomes and instructional strategies
2. Increase knowledge of strategies for incorporating and considering the relative weight of alternative assessment data in the overall determination of student grades
3. Reflect on options for broadening the communication with parents, students, other teachers, and the community about the quality of student work

Authors: Wendy McColskey, SERVE; and Kathy Busick, PREL

Chapter Content

Readings

Grading: Issues and Options .... Page 4

This section provides information about a variety of approaches for incorporating alternative assessments into a grading scheme.

Reporting: Communicating About Student Learning .... Page 13

The focus of this section is on examining assessment as a communication system that engages students, teachers, parents, and others in ongoing conversations about the quality of student work in order to lead to improvement.

References .... Page 19

Activities in This Chapter

Activity 6.1 *Weighty Questions*

Helps to illustrate the importance of developing sound grading practices that reflect the student outcomes that are valued. Time: 90 minutes to 2 hours

Activity 6.2 *How Can We Know They're Learning?*

Engages parents in discussions about the qualities they value in student work and in building criteria that can be used to feed forward into student action. Time: 90 minutes to 2 hours

Related Activities in Other Chapters

Activity 2.5 *Clapping Hands*

Participants play the part of assessors and assessees to explore the consequences of designing alternative assessments in various ways. Time: 75 minutes

Activity 2.8 *Sam's Story: Comprehensive Assessment*

Participants use a succession of assessment information about Sam to draw more and more sophisticated conclusions about his achievement.
Time: 45 minutes

Activity 3.1 *Camping Trip: A Math Problem-Solving Task*

Demonstrates how to develop performance criteria and illustrates how the development of performance criteria can be instructional for both teachers and students. Time: 60 to 90 minutes

Activity 3.2 *Sorting Fish: A Science Task*

This activity is similar to Activity 3.1, but uses a science task instead of a math task. Time: 60 minutes

Activity 4.3 *Performance Criteria—Keys to Success*

Illustrates the characteristics of sound performance criteria. Time: 60 minutes

Sample Assessments Referred to in this Chapter

Sample assessments are used to illustrate the points made in this chapter. All samples, plus an index of descriptive information, are found in Appendix A. The following lists the samples that are referred to in this chapter.

Sample 4.3	*Discovering the Problem of Solid Waste*
Sample 4.5	*Weathercaster's Helper*
Sample 4.10	*Writing in Chemistry*
Sample 5.2	*Assessment of Laboratory Skills in High School Science*
Sample 6.1	*Hawaii Algebra Grading Process*

Using Alternative Assessment in Grading and Reporting—Readings

Grading: Issues and Options

Introduction

Today, the things we value as educational outcomes and what we need as a society have changed. Developing and assessing "big outcomes" (problem-solving, reasoning, communication, critical thinking) are considered at least as important as transmitting and assessing factual knowledge. Numerous efforts are being made across the country to foster instruction that helps students develop these skills. Assessments of these skills are more open-ended and/or performance-based, requiring more subjective judgment in scoring than multiple-choice, factual recall tests. Some teachers may express confusion about how to fit these new methods into traditional percentage-based grading systems. Teachers who work within the constraints of a single grade for six weeks, nine weeks, or a semester need suggestions for how to incorporate information from alternative assessments.

Reporting on student progress can happen in a variety of ways. At primary levels, many schools report student progress in ways other than a single indicator (such as A,B,C,D,F). They might use such things as skills checklists or developmental continuums to track progress. At middle and high school levels, alternatives to report card grades may also be desirable but the traditional report card system may be harder to change. At all levels, report card grades are meant to serve a variety of purposes. They provide feedback to parents and students on how well students are doing in school. They are often intended as motivators (e.g., to impress upon students that the work they do is important and worth their attention and effort). In addition, they recognize good work. On the flip side, they can also be perceived as punishing by those students who receive low grades.

If report card grades are to communicate effectively to parents and students, the basis of their assignment should be clearly articulated. It is not enough just to say that 90 percent or higher is an A if the "90 percent of what" question isn't answered. This section suggests that we clearly define "the what" so that grades communicate effectively.

Effectively Communicating Achievement Status

Assigning a grade involves several steps (Stiggins, 1991):

Step 1: Deciding the student competencies that will be targeted in the course or grade level. What are our goals for students?

Step 2:	Determining the evidence (e.g., student work) that will be gathered to show the extent to which the student competencies have been attained
Step 3:	Deciding how different kinds of evidence will be weighted
Step 4:	Specifying how grades for students will be determined

Steps One Through Three

The first decision—identification of student competencies to be targeted—is the key to the process. If the competencies to be targeted in the class are not clear, the rest of the process is fuzzy. If student goals are clearly stated, not only does instruction become much more focused, choices for assessment strategies and the grading process become much more logical. For example, in a science class, alternative assessments might be incorporated into grades as shown in Figure 6.1. In the sample grading period, conceptual understanding is assessed using paper-and-pencil and interactive computer tests (40 percent of the grade). Alternative assessments such as lab or performance tests, concept maps, projects, and journals are used to assess science process skills, creativity, problem solving, and responsibility (60 percent of the grade).

Figure 6.1
Sample Unit Assessment Package

Student Competencies	Assessment Methods	Weighting in Final Grade
Conceptual understanding	Paper-and-pencil test; interactive computer test	40%
Science process skills	Performance tests	20%
Creative applications of knowledge	Concept maps, creative writing	15%
Research and problem-solving skills	Extended project	15%
Responsible for own learning	Journal or portfolio	10%

Such a chart addresses the first three steps listed above: what student competencies will be developed; what assessment methods will be used (evidence gathered); and how the assessment information will be weighted. Notice that the assessment process becomes more logical when it is organized around student competencies and outcomes. Also notice that this becomes a useful communication tool for students and parents, clarifying the purposes of instruction.

As another example, developers of the Hawaii Algebra Learning Project (Matsumoto, et al., 1994, and **Sample 6.1**—*Hawaii Algebra Grading Process* in **Appendix A**) wanted to develop the student competencies shown in Figure 6.2 and believed that it was important for students to work hard every day and not just study for tests. The following quote from *Assessment Standards for School Mathematics* from the National Council of Teachers of Mathematics articulates the latter concern.

> *Tests are just small tasks...They are used at a point in time, not as a reflection of growth over time. When we grade tests, we are only looking at first drafts because students have not had the opportunity to revise their work due to the time limitations under which the test is administered. The time spent taking a test is a very small part of the time spent using mathematics, so why should the test count a lot?* (1993, p. 132)

Thus, to address the goals of problem solving and communication in mathematics, daily work in the Hawaii project consists of collaborative problem-solving and journal-writing tasks that are weighted more heavily than tests. For example, for a given time period, there might be 26 gradebook entries scored from 0 to 4 (see Figure 6.2). A score of 4 indicates that the student's work meets all the qualifications for excellent work, comparable to an A. Students receive approximately 11 scores on their collaborative group's class presentations on assigned algebra problems. They receive 11 more scores representing the quality of their writing in response to journal prompts such as "How is solving an equation like/unlike solving an inequality?" Finally, to assess conceptual understanding and application of knowledge, students receive one quiz score, one chapter test score, and two lab scores as outlined below.

Figure 6.2
Chapter Assessment Package

Competencies Developed	Assessment Methods	No. of Entries in Gradebook	Estimate of Weight
Math problem solving	Student discussions and group presentations	11	42%
Communication in math	Journal prompts	11	42%
Conceptual understanding	Quizzes, tests	2	8%
Application of knowledge	Labs	2	8%
	Total	26	100%

Hawaii project authors claim that emphasis on daily problem solving results in decreasing fear and anxiety about tests and increasing intrinsic motivation to understand important concepts and to be involved in class discussions. Additionally, students do not memorize for a test and then promptly forget the content after the test, since they are discussing and applying content information daily. Assessment is integrated into the instructional process.

In order to make this work, and have credibility with parents, the teacher must have some way of consistently judging the quality of the student presentations, journal entries, and labs so that an accurate score can be reliably assigned. This is where "rubrics" or "scoring guides" come in. **(Sample 6.1**—*Hawaii Algebra Grading Process* has sample prompts and the rubrics used for scoring journal prompts and group presentations.) Teachers have to practice using these scoring guides so that judgments are consistent across students, days, and assignments. What credibility would such a system have if judgments varied between teachers or differed with the same teacher on different occasions?

Step Four

Figures 6.1 and 6.2 above show how targeted competencies, assessment methods, and weighting (relative importance of competencies) can be matched-up. The last step in determining report card grades is the grade assignment process. That is, how will the final grades be determined based on the information collected in steps one through three? The methods described below are some seen in practice and described in other literature on grading. (See, for example, Frisbie & Waltman, 1992; and Ory & Ryan, 1993.)

Fixed Percent Scale Method

In this approach, scores from each grading component (unit tests, quizzes, performance assessments, observations, homework) are expressed as a percentage. At the end of a grading period, these component grades are weighted and averaged (step three) and assigned a letter grade based on a fixed scale (e.g., 93 to100% = A). This is the same approach currently used by many teachers.

However, this approach can result in several problems in a classroom where the teacher is teaching to complex instructional objectives and assessing these "big outcomes" with alternative methods.

1. Knowing that 93 percent *always* means an A may lead to restricting what the test covers.

> *If the teacher believes there should be some A grades, a 20 point test must be easy enough so that some students will score 19 or higher; otherwise there will be no A grades. This circumstance creates two major problems*

> *for the teacher as assessment developer. First, it requires that assessment tasks be chosen more for their anticipated easiness than for their content representativeness. As a result, there may be an over-representation of easy concepts and ideas, an overemphasis of facts and knowledge, and an under representation of tasks that require higher-order thinking skills.* (Frisbie & Waltman, 1992, p. 40)

Additionally, a student score of 80 percent might represent different performances on two different tests. It may be that one test assesses higher-order thinking and the 80 percent actually represents a better performance than 80 percent of a short-answer knowledge test.

2. Secondly, scoring guides (rubrics) for alternative assessments describe the level of a student's performance on a task and, thus, are essentially descriptive. Converting the descriptive score to a percent score makes no sense. If five stars are the maximum possible and a restaurant receives a three star rating, it would make no sense to convert the 3 to a 60 percent, and judge the restaurant a failure. Three stars may actually mean "ok" or even "good."

   For example, consider **Sample 4.3**—*Discovering the Problem of Solid Waste* or **Sample 4.5**—*Weathercaster's Helper* in Appendix A. Both are scored on a variety of dimensions using a point system. In **Sample 4.5**, the total points possible is 11 and in **Sample 4.3**, the total possible is 21 points. The points are assigned based on how well students' responses match the desired responses. The relevant information for the student rests with where points were lost not with the percent of total points obtained—75 percent might actually indicate good (B) performance.

   Also consider **Sample 4.10**—*Writing in Chemistry* which uses an analytic rubric to score responses to a chemistry writing task. Each essay is scored from 0 to 5 on six dimensions (general impression of content quality, prior knowledge, number of principles or concepts, argumentation, text, and misconceptions). The information to the student is contained in these individual scores. The profiles of students' scores tells them their strengths and weaknesses (e.g., strong on "argumentation," weak on "number of principles"). Deriving a percent score would only confuse the message to students. A student who received all 3's would fail the task with a 60 percent, using a percent scoring system, whereas 3's may represent minimally acceptable performance (C work) in the teacher's judgment.

For the above reasons, a percent averaging system seems to be incompatible with alternative assessments. Other options are described below.

Total Point Method

Some teachers have students earn points throughout a grading period, then assign grades based on the point total at the end of the period. For example, the teacher may determine that for a particular grading period points will be accumulated as shown in Figure 6.3. Other examples: **Sample 4.3**—*Discovering the Problem of Solid Waste* shows a performance test scored with a point system which *might* represent 11 of the 40 points for conceptual understanding. **Sample 5.2**—*Assessment of Laboratory Skills in High School Science* is an example of how 35 points on a lab test *might* be accumulated to assess science process skills. Finally, to assess communication skills, a sixth-grade journal might be checked for completeness of entries each week with a maximum of 10 points per week.

Figure 6.3
Sample Point Assignments for a Grading Period

Objectives	**Assessment Methods**	**Total Points Possible**
Conceptual understanding	Paper-and-pencil tests	40
	Interactive computer test	50
Science process skills	One lab	35
	One demonstration summary	15
Creative applications	Two tasks at 10 points	20
Plan/implement/share independent learning	One project	50
Communication skills	Journal	60
	Total	270

Point values for the assessment tasks and the proposed cutoffs for grade assignments should be determined prior to starting the grading period. For example, in Figure 6.3 students might be told that 235 or higher would earn an A. The decision about where to draw the line for an A, B, C, D, or F depends on the teacher's standards and expectations about what students should accomplish, not necessarily on "90%=A."

Figure 6.4 shows how totaling points might be a more equitable way to assign grades than deriving percents and then averaging, even in a system where policies require a scale such as A=93 to 100%, B=85 to 92%, C=77 to 84%, D=70 to 76%, F=< 70%.

Figure 6.4

Assignment #	Points Received	Converted to Percent
1	9 out of 11	80%
2	45 out of 50	90%
3	2 out of 5	40%
4	14 out of 20	70%
5	40out of 50	90%
	110 out of 136	AVG.
	81% = C	74% = D

The point system approach may fit well in classrooms where teachers use point system rubrics. For example, students may be asked to write up science demonstrations several times per grading period. Each write-up may be worth 20 points (5 points each for a solid hypothesis, method, results, and conclusion statement). A science project may be worth 100 points with points awarded for a variety of criteria. Another point system might be used for demonstrations of effective cooperative group skills.

This approach may also work well for teachers who like to provide incentives for students to complete their work in addition to grading work on its quality. For example, the total points awarded over a grading period are a combination of points for completing work that is designed to be instrumental in developing a particular competency in addition to points awarded for the level of performance on unit and final assessments.

For example, one middle school science teacher has her students keep a "science points score card" which is given out at the beginning of each class period with opportunities during the class to earn points for completed work. Over a grading period, students have the opportunity to earn a large number of points (e.g., 600) both for completing work and for graded assignments or tasks. This approach may be particularly effective in the middle grades to decrease "test anxiety" and fears of failure and to keep students involved and feeling successful. Accumulating points for completing work, when the work has instrumental, instructional value increases the likelihood that students will feel successful and, thus, stay involved in the learning.

Standards-Based Method

In this approach, teachers develop scoring guidelines for assignments to be included as components of the final grade. In the Hawaii Algebra Learning Project **(Sample 6.1)**, all student assignments are scored on a five point scale. The scores received are descriptions of mastery related to the particular skill area. For example, an A or 4 on a journal entry means that students communicated well, supported statements with examples, and extended the answer by suggesting an exploration, creating a new question. "Generalized" rubrics are developed for each student outcome (problem solving, communication through journals, etc.).

Figure 6.5 shows sample results for a particular student.

Figure 6.5
Sample Student Record for a Unit

Assessments Used		**Score**					**Weighting**
		0	1	2	3	4	
Problem-solving tasks	#1			X			44%
	#2			X			
	#3				X		
	#4					X	
	#5					X	
	#6					X	
	#7				X		
Communication tasks	#1			X			44%
	#2			X			
	#3					X	
	#4					X	
	#5					X	
	#6					X	
	#7					X	
Tests/application tasks	#1					X	12%
	#2					X	

If the student needed to know how he or she had performed for this period of time, the scores could be averaged. In this case, the average is 3.4 [((4x2) + (2x3) + (10x4))/16], which could be converted to a percentage score for school purposes, if needed. Or some teachers may weigh later work more heavily than early work because it may be more reflective of what the student can do. The early scores of "2" may reflect performance prior to understanding the criteria for good work. (See the next section for additional discussion of this option.)

To inform parents, samples of work representing the four levels of quality could be kept so that parents could see how their child's work compared to higher or lower levels of work.

This standards-based approach is most consistent with the current national climate of developing performance standards and benchmarks. It may be time-consuming, at least at first, because of the need to define levels of achievement on tasks and help students understand the criteria for particular skills. However, as discussed in Chapter 3, such efforts are not empty exercises designed merely to grade students; they are also necessary components to improve instruction and assist students to take control of their own learning. For example, in the Hawaii Algebra Learning Project, students participate in the development of the criteria for group presentations and journal prompt responses. These are "generalized criteria" and can be applied across different kinds of problems solved and journal prompts. Because they are repetitively used, students internalize them and constantly assess their work against them.

Should grades include performances from early in the grading period or just those at the end?

As Figure 6.5 demonstrates, students may perform more poorly on the beginning assignments than they do on later assignments assessing the same skill. For example, as students continue to write answers to journal prompts, one would hope that they would improve their mathematics communication skills. Early scores for mathematics communication may not present as accurate a picture of student ability to communicate in mathematics as scores at the end of the grading period. So, it could well be argued that if parents are to be provided an accurate picture of their child's achievement level, only the mathematics communication scores for the journal entries at the end of the grading period should be considered. In other words, who cares if skill level is low at the beginning, as long as skills are good at the end? (But notice that this approach requires a good description of "adequate skill" so that the teacher knows when the student is consistently performing at that level; and so we return to the necessity for good quality rubrics and training in their use. Also note that this approach might work better for "big" student outcomes such as problem solving, critical thinking, and communication and group collaboration than for content knowledge.)

Should opportunities for revisions of work be allowed in certain skill areas (e.g., communication)?

In the development of certain skills (e.g., critiquing science articles, writing research reports), students learn from the opportunity to revise products based on specific comments from the teacher. Grading these assignments the first time they are submitted may not contribute to the improvement of students' skills. In these situations, students might submit "finished" drafts to which teachers respond with suggestions for revision. Students who complete the revisions might earn more points for the assignment than students who choose not to make revisions.

In the Hawaii Algebra Learning Project, students can receive a maximum score of 4 for their group presentation of the problem they solved. If the group or any individual within the group is unhappy with the score they received, they have the option to redo the discussion of the problem on paper and turn it in for an improved score.

By encouraging revisions, the message to students is that improvement is valued. Also, practicing revision will tend to ensure that succeeding performances will be better, especially if the criteria for success are crystal clear.

Should student choice (e.g., goal-setting and assembling evidence of their own learning) be part of the grading process?

Another choice facing teachers as students become active partners in the learning process is the degree to which responsibility for proving the attainment of target outcomes is handed to students. Should a grade be a negotiated understanding between the teacher and student about the quality of the student's work based on evidence provided by the student?

Perhaps the targeted competencies to be developed in the course are explained to students. Students might be required to develop a portfolio of the work that best demonstrates their achievement and write a paper arguing for the grade they think they should earn. In this way, the grading process becomes part of the effort to develop students' capacities for self-assessment.

Reporting: Communicating About Student Learning

The expanding interest in alternative assessment has major implications for what we have called "reporting." Teachers who value the insights and student engagement that come from using alternative assessments such as open-ended performance tasks, exhibitions, and portfolios often worry about how to "report" student learning using such assessments. The word "reporting" itself indicates a view of the communication between teacher and students, teacher and parents, or school and community as one-way. Some mathematics and science educators are calling for us to reframe our definition of assessment to incorporate the essential role that it has as a communication system—for opening up conversations between students and their teachers, between students and parents, teachers and parents, and among students.

Engaging others in ongoing conversations based on rich, detailed portraits of student learning is central to the view of assessment as a communication system, a system for making meaning of children's learning and for taking action to improve teaching, curriculum, and learning. Thus, questions about "reporting" actually may be more about changing the way we communicate. (Please note that the grading suggestions in the previous section have the effect of changing the way we communicate progress to students and parents. They also have the effect of including students and parents in the assessment process.)

Because we know that assessment information communicates powerfully what is valued within a classroom, community, region, or nation, we must examine much more closely what, and how well, our current reporting systems communicate what is intended. Schools and communities need to deal with questions such as:

- What intelligences, understandings, capabilities, and habits of mind are valued locally?
- How does what is valued locally relate to standards of excellence for all children?
- Do our present assessments give us (students, parents, teachers, community) insights into the things that we value?
- What do we want our assessment to do for us?
- What are our assessment questions about student learning?
- How will action follow from ongoing communication about learning?
- Who is responsible for communicating about learning?
- Who can take the lead?
- What decisions rest on assessment information?
- How clear is our current picture of student learning?

Until we answer these questions, it will be difficult to develop a satisfactory reporting system. A first step is to work through a shared view of the purposes for assessment. Figure 6.6 shows an example of an initial brainstorm.

Figure 6.6
Brainstorm: Purposes for Assessment

What do we want assessments to do for **students**?

- Reveal their true capabilities
- Enable and encourage self-assessment; evaluate their own progress
- Prompt critical thinking
- Help them become responsible for their own learning
- Emphasize what's important
- Reflect and evaluate their own progress and learning

What do we want assessments to do for **teachers**?

- Deepen the curriculum
- Be consistent with other standards
- Become better *listeners*; know their students well
- Self-assess what they're doing; become deep thinkers
- Use assessment information to improve instruction

What do we want assessments to do for **family**?

- Create clearer perceptions of children and their teachers
- Strengthen communication
- Enable the family to have a greater role in schooling

More thoughts:

- Assessment should contribute to children's learning
- Teachers must be full partners in making the changes (instruction, curriculum, and assessment)
- Professional development—any assessments imposed without training and technical assistance for teachers are doomed

These and many more questions can guide the building and refining of a communication system that engages learners, teachers, families, and communities in conversations that are worth the time spent on them, that feed forward into action rather than back to simply mark where a child's work was at a particular point in time.

Options for Communicating About Student Learning

Frequently the issue of reporting is framed as redesigning report cards. As an alternative or supplement to current report cards, a number of schools are implementing reporting systems that more fully match their beliefs about the purposes of assessment in the overall picture of school learning. Some common characteristics of such alternatives are an emphasis on communicating about complex skills and habits of mind; multiple intelligences and collaborative work, using narratives or other alternatives to letter or number grades; and often, reporting by themes or major concepts across subject areas. As with most innovations, the responses of schools and teachers to rethinking and redesigning communication about student learning can spread widely across a continuum—ranging from efforts to provide some additions to the existing categories reported, to wholly narrative and descriptive summaries, to student-led conferences that take the place of traditional report cards.

The following ways to demonstrate student achievement are examples from among many possibilities. They represent several options most unlike traditional report cards. Further information and examples can be located through the Griffin Center for Human Development*. The Griffin Center has been tracking current trends in report cards for a number of years.

Portfolios

Schools like Central Park East Secondary School in New York City use portfolios as demonstrations of learning. Students prepare 14 separate portfolios over an extended period of time. Part of their responsibility is to communicate about their learning through the portfolios and the required "defense" of several portfolios before a panel that includes a long-term student advisor, an adult selected by the student, another student, and someone with extensive knowledge of the content of the portfolio. Rubrics for scoring portfolios spell out expectations and have included such things as viewpoint (encompasses wide knowledge base and is focused), connections (the whole is greater than the sum of the parts), evidence (credible and convincing), voice (engaging), and conventions (intelligible). Students determine their readiness to present their portfolios and get very focused information on what to do differently when the portfolio is not yet considered as having met expectations.

In Hawaii, an elementary school is seeking to use portfolios as part of a reporting system that links to an overall vision for students who are becoming self-confident risk takers and creative problem solvers, collaborative and capable in multiple dimensions. Parents, teachers, and students struggle to define each vision element. Primary teachers have invited parents to respond to several report options. Figure 6.7 is an example of part of a first grade report that students helped create.

---

* Located at 15 South Fair Street, Guilford, Connecticut 06137

Figure 6.7
Sample First Grade Report Card

D = Developing **P = Practicing** **I = Independent Application**	**Student**	**Teacher**	**Parent**
Creative Problem Solver Looks for different ways to solve a problem Knows where to find information			
Self-Confident Risk Taker Willing to participate in activities and discussions Willing to try new things Willing to try independently			
Collaborative Takes care of property Cooperates in a group Uses kind words Listens to others' ideas Finishes work on time			
Well-Rounded Willing to try a variety of activities			

Narratives

Schools such as Crow Island School in Winnetka, Illinois, use narratives as well as portfolio evenings to engage students and parents in thinking and communicating about learning. An increasing number of schools that have begun using narrative descriptions of learning are incorporating technology into their reporting systems. Juneau School District in Alaska is using computers to support teacher narratives.

Using a Continuum

Teachers in some schools have worked to describe the common characteristics of readers, writers, etc. at various developmental stages. Reports using this option acknowledge that children's learning often spreads out across several "levels" rather than concentrating at a single point. Teachers might check or highlight the skills, behaviors, and knowledge that students currently exhibit and use this as an indicator of developmental level.

Multiple Sources

FIRST CLASS, an innovative middle school program in Rochester, New York, uses many forms of evidence of student learning in an ongoing reporting system, including report cards, interim interviews, written reflections and video/audiotaping, portfolios, conferences with families, presentations, exhibits, and standardized testing.

The report card for FIRST CLASS reports in four major areas: (1) engagement/class participation and involvement; (2) collaboration, cooperative learning groups, and working relationships with others; (3) independence and self-direction; and (4) performance/quality of work.

Student-Involved Conferences

Schools in several districts in British Columbia have been using student-led and three-way conferences as alternatives to formal report cards. Three-way conferences take place when a child meets with his/her parents and teacher to demonstrate, show, and/or tell what has been learned. The meeting is "chaired" by the child. The child's learning strengths and areas of concern are discussed. Learning goals are set with the child, parent, and teacher. The teacher is available for questions, but the student has major responsibility for communicating about learning. As currently designed, brief written summaries of such conferences are permitted in place of more formal report cards several times a year. This option is part of a much larger system of ongoing communication among student, family, and teacher.

Conclusion

A central concern in any examination of reporting is the need for a good match among the learning that is desired, the forms of assessment used to gather rich images of learning, and the communication system for conversations about learning that lead to action and growth. To gain insights into the appropriateness of various reporting options, ask yourself and others what the report card or report system communicates about the things that are valued within the school or system. What things are reported directly? What are not? How consistent is the reporting system with valued outcomes for students? Is information communicated in a way that is clear and promotes understanding? If so, then the reporting system works.

References

Davies, A., Cameron, C., Politano, C. et al. (1992). *Together is better: Collaborative assessment and reporting.* Winnipeg, Canada: Pequis Publishing.

Frisbie, D. & Waltman, K. (1992). Developing a personal grading plan. In *Instructional topics in educational measurement.* Washington, DC: National Council on Measurement in Education.

Hebert, E. (1992, May). Portfolios invite reflection-from students and staff. *Educational Leadership,* pp. 58-61.

Matsumoto, A., Dougherty, B., Wada, L., & Rachlin, S. (1994). *Algebra I: A process approach teacher's guide.* Honolulu: University of Hawaii, CRDG.

National Council of Teachers of Mathematics Assessment Standards Working Groups. (1993). *Assessment standards for school mathematics: Working draft.* Reston, VA: Author.

New York Board of Education. (1991 revision). *Senior institute handbook.* New York: New York Board of Education, Central Park East Secondary School.

Ory, J. & Ryan, K. (1993). *Tips for improving testing and grading.* Newbury Park, CA: Sage Publications, Inc.

Stiggins, R. (1991). *A practical guide for developing sound grading practices.* Portland, OR: Northwest Regional Educational Laboratory.

Sundberg, N. & Daniel, S. *First class.* Rochester, NY: Frederick Douglass Middle School, Rochester City School District.

ACTIVITIES

Activity 6.1

Weighty Questions

Purposes:

1. To discuss the importance of developing sound grading practices that reflect valued student outcomes
2. To provide practical examples of how to incorporate alternative assessments into report card grades
3. To examine current testing and grading practices in relation to desired outcomes for students

Rationale:

Teachers, especially at higher grade levels, may see the need for alternative assessments such as journals, portfolios, projects, etc., but have a difficult time envisioning how to make them fit with district demands that the assignment of grades be based on a percent system (e.g., 93 to 100%=A). This activity supports teachers in articulating the choices they have made in their current testing and grading processes and making decisions about what they would like to change.

Materials:

- **Sample 6.1**—*Hawaii Algebra Grading Process*
- **Handout A6.1,H1**—*What Outcomes Do Should Grades Assess?*
- Chart paper, markers
- Chapter content used as a handout: Chapter 6

Time Required:

90 minutes to 2 hours

Facilitator's Notes:

A. Basic Activity

This activity should probably follow a series of activities that have familiarized participants with alternative assessment. It assumes that they will be familiar with a variety of alternative assessments (Chapter 4) and the idea of matching assessments to valued outcomes and using assessments as an instructional tool (Chapter 3). The reading material in Chapter 6 could be distributed and read prior to this session.

1. Review the kinds of decisions (outlined in Chapter 6) that are made in giving report card grades, perhaps using **Sample 6.1**—*Hawaii Algebra Grading Process* as an illustration.

2. Introduce the activity by telling participants that they will be discussing the decisions about grading they have made. Pass out **Handout A6.1,H1**—*What Outcomes Do/Should Grades Assess*? Ask participants to *individually* complete the first three columns for a course or subject area they teach, thinking about the current or last grading period. For example, a science teacher might list quizzes and tests under "assessment methods," indicate that the "knowledge/skills assessed" by the tests used were recall, conceptual understanding, and application, and record that this assessment method accounted for 60 percent of a student's report card grade. They might list projects as the next assessment method used, and under "knowledge/skills assessed" enter plan, research, synthesize, create a product, present information, develop conceptual understanding, and defend a point of view. They might say projects counted as 20 percent of the final grade. And so on.

3. After participants have had time to complete the information individually, put participants in groups of four and ask them to share their responses in the three columns. Ask each group to try to come to some kind of consensus about the desired outcomes and weighting for the course/grade they teach. They should complete the last two columns to report out to the whole group. The reporting might prompt a discussion of whether or not teachers in a school should try to come to some kind of agreement about important outcomes to be assessed in the grading process.

4. Finally, ask each group to refer to the pages in Chapter 6 describing how grades might be assigned (percent averaging, total point, standard-based, other options) and list the methods used by group members on chart paper. The methods used could be other than described. Groups should report back to the whole group with discussions of any questions, helpful suggestions, etc.

B. Extensions

1. As a follow-up activity, teachers could be asked to bring a test they have recently used and work in groups to list the kinds of skills assessed by each item (e.g., recall, analysis, comparison, inference, evaluation). The point of this analysis would be to (1) find out if the tests assess the kinds of skills or competencies that the faculty value, and (2) to continue conversations among teachers in a school about what outcomes are valued and how assessments might be brought more in line with valued outcomes.

2. They might also be asked to discuss:

    a. The extent to which they modify grades for student effort, generating a list of the pluses and minuses of doing so

    b. The extent to which they modify grades in consideration of student ability and the pluses and minuses of doing so

What Outcomes Do/Should Grades Assess?

List assessment methods reflected in gradebook (e.g., paper and pencil tests/quizzes, homework, journals portfolios, projects, performance tasks)	Add knowledge/skills (outcomes) assessed by each method	Estimate current weighting (% of total grade)	Reconsider outcomes: List knowledge/skills that should be assessed	Enter desired weighting for each outcome area (% of total grade)
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				

A6.1,H1

Activity 6.2

How Can We Know They're Learning?

Purposes:

1. To use samples of student work as a way to examine what we value
2. To broaden understanding of the importance of clear criteria for quality assessment
3. To contribute to the criteria that can be used to assess their children's work
4. To broaden understanding of assessment techniques

Rationale:

This activity, an adaptation of **Activity 3.1**—*Camping Trip* and **Activity 3.2**—*Sorting Fish*, is intended for parents and is centered on the question "How can we know they're learning?" As schools and classrooms expand the use of alternative and performance assessments, it is vitally important that parents are invited to become an integral part of the conversations about student learning. In the past, such conversations have focused on brief, often numeric, summary information that does not provide parents with insights into the richness of learning—the expanding knowledge, skills, and capabilities that children are gaining.

Using student responses from **Activity 3.1** or **Activity 3.2** (or any set of 8-10 student responses to a common task), this activity is an opportunity for parents and family members to expand their knowledge of alternative assessment and to become involved more fully in rich and valuable conversations about children's learning. In addition, this activity includes responses from a mathematics educator to enable parents to look inside the reasoning that teachers use when assessing student work.

This activity is intended as an introductory session with parents, community members, and others. It has been used as part of a series of sessions with parents over time that builds a common understanding using "neutral" student work, and then gradually moves to examination of anonymous samples from the same school, and finally to examination of the work of their own children once the language of assessment has been established, and agreement on qualities to look for in the work are clear. The sessions often include teachers and students.

Materials:

- Overhead projector, screen, blank transparencies, transparency pens, or chart paper and pens
- Eight to 10 responses representing different levels of sophistication; for example, the set of student answers to problem solving-exercises in **Activity 3.1**—*Camping Trip: A Math Problem-Solving Task* (other student responses, e.g., **Activity 3.2**—*Sorting Fish: A Science Task* also work)
- **Overhead A6.2,O2**—*Holistic and Analytical Trait Scoring,* **Overhead A6.2,O1**—*Assessment Is About the Exchange of Information*; **Handout A6.2,H1**—*The Redwoods* and **Handout A6.2,H2**—*Fox*
- **Sample 4.2**—*Mathematics Problem Solving*, and **Handout A6.2,H3**—*Waialae's Draft General Problem-Solving Guide* (one set per group of two-three); other generalized math problem solving scoring guides could also be used; (for example, **Sample 4.7**—*Performance Assessment in Math (Alberta)* or **Sample 5.7**—*Mathematics Assessment)*

Time Required:

90 minutes to 2 hours

Facilitator's Notes:

A. Basic Activity

1. Open the session with **Overhead A6.2,O1**—*Assessment Is About* and use it to underline the purposes of the activity.
2. Provide a mini-practice that emphasizes how to describe the qualities of student work using overhead samples of student writing. (Writing is chosen because it is an area where most parents feel comfortable with their ability to comment about quality.) Begin with the piece that represents a relatively strong response (e.g., **Handout A6.2,H2**—*Fox*). Ask the whole group to review the piece and identify its strengths; its key qualities. Record their responses on one side of a blank transparency.

   Then discuss **Handout A6.2,H1**—*Redwoods*, a beginning response, and ask groups to identify what they see as the key qualities of this response, recording their answers on the other side of the transparency.

The purpose is not to come to complete agreement, but to provide an opportunity for parents to build on their own experience and contribute to a discussion of the qualities found in student work.

3. Have participants move to groups of two or three. You might, for example, have parents add the ages of their children/grandchildren and find two other people close to their own total.

4. Bridge to math problem solving by noting that teachers, too, are engaging in conversations about the quality of student work and that they are especially looking at the big ideas that are central to learning for all children—like problem solving. (If there is a school vision in place, note the connections between the vision and students' becoming effective problem solvers.) If one purpose of this activity is to connect assessment and national standards, identify the NCTM mathematics problem-solving standard as one which cuts across ages and grades.

5. Parents should put the student solutions (from **Activity 3.1**—*Camping Trip*) into three piles based on the "quality" of the solution. Describe this task in ways that cue what to do. For example you can frame the task as dividing the work into levels of sophistication, using language such as "beginning work", "developing work that shows both strengths and weaknesses", and "strong work." Or it might be helpful to cast participants as those responsible for grading: "Put papers into stacks by what grade you would give them." (Note that using grading as the prompt often results in questions and discussion of current grading practices; this is great when it's intended, but needs to be anticipated so that the activity doesn't lose its focus.)

   As participants sort the solutions, they should write down the basis on which they placed papers in each stack. *You thought the papers in this stack were better than the papers in this other stack. Why? What features of the work did you like? Not like? What makes the work in this stack different from the work in that stack?* One strategy is to designate a person to be the observer/recorder for each group whose responsibility is to capture the evaluative language used, and the reasons cited for placement. (When there are teachers present, this is often their role so that parents do not defer to their judgment and limit their own discussion.)

   The goal is not to score or "grade" all the papers, but to articulate reasons. Therefore, participants only continue sorting and discussing until all their reasons are listed. Remind each group of this as they work.

6. Put up a clean overhead transparency (or use chart paper) with three columns labeled "Strong, developing, and beginning" or "high, middle, and low." (Use whatever titles will appeal to the group.) Have groups tell you the placement of

each paper. Start a fourth column with a "?" heading. If there is a disagreement about placement, write the paper number under the "?" heading.

There is no "right" answer for this sorting. The goal is for participants to articulate what they value, come to consensus, and be able to justify their judgments using consistent terminology.

7. List all the criteria the parents gave for their sorting.

8. Try to resolve placement of the "?" papers. Have the various groups describe their rationale for placement. Add their reasons to the criteria list. The goal is to have a rich and elaborated list of criteria that cover many features of what is valued.

   Depending on time, you can use disagreements as an opportunity for parents to clarify and articulate what they saw in particular papers. This is also an opportunity to point out that while it is possible and important for educators to come to agreement on key qualities of student problem solving, there are also differences among what individual teachers value. The point can be made that making what is valued overt and clear to students is critical to quality assessment.

9. Using Oregon's four-trait model, **Sample 4.2**—*Mathematics Problem Solving* or another scoring guide (e.g., **Sample 4.7**—*Performance Assessment in Math* or **Sample 5.7**—*Mathematics Assessment*), ask parents to see the connections between their list and the sample scoring guides. For most groups, their own careful look at student work resulted in the identification of quite similar key dimensions or traits. The point is not to simply underline how parents' ideas echo the scoring guides of others, but to understand that the process they have used is one that professional educators also engage in as they seek to clearly define assessment targets.

   To emphasize the common ideas as well as local features of models developing among schools around the country, have parents take a look at **Handout A6.2,H3**—*Waialae's Draft General Problem-Solving Guide*. This piece is undergoing revision as parents, students, teachers and community advisors integrate their ideas and implement assessments embedded in thematic units across the school.

10. Close with oral or written parent reflections. Some prompts might be:

    - What insights did you gain?
    - What questions remain?

B. Extension

1. The session can be extended to provide awareness of the benefits and drawbacks of different kinds of scoring. You might pose the question, *Did you see some quality or qualities in a work that made you want to place it in one stack, and then other qualities that seemed to place it in a different stack? What were they? How do you feel about the forced choice of one stack or the other?* Note that the forced choice that they used echoes holistic scoring in which the overall impression of the work guides scoring.

   Use **Overhead 6.2,O2**—*Holistic vs. Analytical Trait Scoring* to clarify the distinction between holistic and analytical trait scoring. Ask the groups to discuss when each kind of scoring would be useful. What are the positives for parents and students if student work is scored analytically? What are the positives for parents and students if student work is scored holistically? What are the negatives of each?

   This can also lead into discussion of what kinds of reporting helps parents understand their children's learning and gives them some pointers about what's worth praising in their child's work and what areas need more attention.

2. Another extension is to try it out: If time, take one or two new solutions to the same problem, or a sample of student responses to another problem solving task, and try to score it analytically using the four traits in **Sample 4.2**—*Mathematics Problem Solving* as modified and clarified by group discussion. As before, parents should justify their judgments using phrases or descriptors from the scoring guide.

The Redwoods

Last year, we went on a vacation and we had a wonderful time. The weather was sunny and warm and there was lots to do, so we were never bored.

My parents visited friends and took pictures for their friends back home. My brother and I swam and also hiked in the woods. When we got tired of that, we just ate and had a wonderful time.

It was exciting and fun to be together as a family and to do things together. I love my family and this is a time that I will remember for a long time. I hope we will go back again next year for more fun and an even better time than we had this year.

A6.2,H1

FOX

I don't get along with people to good, and sometimes I am alone for a long time. When I am alone, I like to walk to forests and places where only me and the animals are. My best friend is God, but when I don't believe he's around sometime's, my dog stands in. We do every thing together. Hunt, fish, walk, eat and sleep together. My dog's name is Fox, 'cause he looks like an Arctic Fox. Fox and I used to live in this house with a pond behind. That pond was our property. The only thing allowed on it (that we allowed) was ducks & fish. If another person or dog would even look like going near that place, Fox and I would run them off in a frenzy. There was a lot of rocks around, so I would build forts and traps for any body even daring to come near. The pond had a bridge that was shaded by willows, so on a hot day me and Fox would sit on that bridge & soak our feet, well, I would soak my feet, Fox just kinda jumped in.

At night, the pond was alive with frogs, so I would invite this kid over, (he was a guy like me) and catch frogs. After we had a couple each, we would pick the best looking one out of our group and race them. The winner gets the other guys frog.

In the winter, the pond would freeze over, and I got my iceskates out. The pond was now an ice skating rink. Fox would chase me as I went round & round the pond.

After about a year, I was riding my bike patroling the area around the pond. With Fox at my side, I raced downhill toward the pond. I tried to stop, but my back tire went into a skid. I went face first into murky, shadowy waters. When I went down, a minute later I felt something pull on my shirt, I grabbed it, not knowing what to think, when I hit the surface, I saw that it was Fox, pulling on my shirt as if he was trying to save me. He was to little to save me if I was really drowning, but it was the thought that counts, I owe him one.

Another year passed. One day my mom got home from the store, and she bought me a rubber raft. It was just a cheap one, but it was mine. I blew it up with a tire pump. It was just the right size for me & Fox. Out of respect for Fox, I named it the USS Fox and christened it right in the pond.

On sunny days, I would take the raft out & lay in the sun with Fox on my legs. One day, when I was asleep in the raft, the wind blew pretty hard and blew my raft right into a bunch of sticks and rocks, the USS Fox was given a sad salute, and then was no more.

Another year passed, and this would be our last year by the pond. I admired and respected that pond more than I ever did that year. But, at long last, all good things must come to an end, we moved to another town. Fox & I still visit the pond, but it'll never be like them 3 years when she was mine.

Waialae School
Preliminary Draft of Standards and Indicators for Creative Problem Solving

Draft #3

Definition of Creative Problem Solving

Creative problem solving includes the ability to understand the problem, build solutions and follow through, and produce an outcome which addresses or solves the problem. It includes an element of creativity which requires that children extend beyond their own limit in the problem solving process or outcome and involves the ability to see a multitude of possibilities. Creative problem solving is developed through continuous opportunities to perform authentic tasks and involves children being able to construct meaning or find connections within the larger context of real world situations/problems.

Standard #1: Understands Problem

- Distinguishes between relevant and irrelevant information and factors in the problem
- Links experience and prior knowledge to problem
- States the problem in a question or series of questions
- Reflects on possible multiple approaches to arrive at outcomes and/or possible multiple outcomes for the problem
- Relates problem to real world situations and contexts
- Makes meaning of the problem in personal, social or community sense
- Interprets problems from other points of view and multiple perspectives

Standard #2: Building Solutions and Following Through

- Making a Plan
    - Draws on and applies past knowledge
    - Identifies information and resources
    - Evaluates multiple approaches
    - Selects appropriate, feasible, efficient, simplest solution strategy
    - Describes plan of action

- Following Through
    - Accesses and gathers appropriate materials and data
    - Takes risks by trying ideas that are new or original for the child
    - Evaluates appropriateness of strategy
    - Adjusts strategy based on evaluation
    - Explains and describes what worked and did not work during process to create final product

A6.2,H3

Standard #3: Outcomes

- Originality
    - Product is a stretch of child's limits
    - Product is unique for child
    - Uses alternative resources/materials

- Quality
    - Product is clear and coherent
    - Product is accurate
    - Product is refined and polished
    - Product is organized
    - Demonstrates effort
    - Demonstrates persistence

- Appropriateness
    - Product/process is related/connected to problem or task

- Reflection
    - Evaluates product and makes suggestions for improvement

Standard #4: Creativity

- Flexibility in Thinking
    - Generates multiple approaches

- Transfer in Thinking
    - Applies approaches to problems to new situations and contexts
    - Communicates verbally and nonverbally about application of concepts to new or real world situation

- Originality
    - Demonstrates divergent thoughts and approaches which are different from past thoughts and practices for that child
    - Demonstrates divergent thoughts and approaches which are unique for all
    - Reflects on the uniqueness of thoughts and approaches

A6.2,H3

Assessment is about the exchange of information

Assessment is not primarily about the collection of information—but rather about its use in communicating about learning.

Information flows both ways. We have to respond to what the student creates and how s/he responds.

It's NOT simply a tool for saying, "Bad luck, you got it wrong."

David Clarke
September 28, 1993

A6.2,O1

LABORATORY
NETWORK PROGRAM

Holistic Scoring

Scoring that responds to a student's work as a whole; assigning a single score or description of quality based on clearly defined criteria.

Analytical Scoring

Scoring that responds separately to each of the key qualities or dimensions of the student's work; assigning individual scores or quality descriptions for each dimension.

A6.2,O2

LABORATORY NETWORK PROGRAM

Chapter 7

Designing Effective Professional Development

What's in this Chapter

The activities included in the various chapters of the *Toolkit* have been developed to enable facilitators to design appropriate professional development opportunities on the topic of alternative assessment for a variety of audiences and purposes. The goal of this chapter of the *Toolkit* is to provide users with a brief review of professional development strategies so that effective professional development using the material in the *Toolkit* can be designed. In this chapter, the term *professional development* is used broadly and is not meant to refer only to classroom teachers. Although the purpose of individual professional development activities and opportunities will vary, this chapter assumes that the ultimate result of professional growth and development will be the improvement of student achievement.

Chapter Goals

1. Provide an overview of the literature on how to plan effective professional development
2. Provide sample training agendas for short sessions of various lengths
3. Illustrate longer professional development approaches using vignettes

Chapter Content

Readings

Characteristics of Effective Professional Development ....Page 3

The chapter includes a discussion of:

361

1. What facilitators need to know about their clients in order to plan effective professional development

2. Lessons from adult learning theory that relate to professional development

3. Various models for professional development

4. Specific research on the training model

5. The changing paradigm of professional development

Sample Short-Term Training Plans.. Page 10

These plans are based on actual training sessions conducted by LNP members. All make extensive use of *Toolkit* activities, chapters, and samples. Some assessment expertise is required of the facilitator.

Long-Term Professional Development Vignettes .... Page 14

This section includes descriptions of actual ongoing professional development sequences, extending over a year or more. Extensive assessment expertise is required of the facilitator.

Other Resources

Additional Readings and Resources .... Page 22

The bibliography suggests additional readings on the topics covered in this chapter.

Workshop/Training Evaluation Forms... Page 24

Two evaluation forms are included—presenter and participant. The presenter form is intended (1) for the facilitator to reflect about what aspects of the presentation worked and which need to be revised, and (2) to provide feedback to the *Toolkit* developers so that future editions can be continually improved. The participant evaluation form provides feedback to the presenter on the quality and usefulness of the training session.

362

Designing Effective Professional Development—Readings

Characteristics of Effective Professional Development

Professional Development and Systemic Change

Changes in assessment approaches necessitate corresponding changes in instructional approaches and vice versa. Professional development efforts that focus only on assessment without concern for pedagogical approaches or for the myriad of interconnections between associated aspects of the educational system do teachers a disservice. (See **Appendix B** for a discussion of how this *Toolkit* fits into the larger systemic reform effort.)

There are many resources that can assist in planning and delivering both short- and long-term professional development programs of high quality. Some of these are listed on page 21. In addition, a separate taskforce of the Regional Educational Laboratory Network Program has developed a complementary set of materials, *Facilitating Systemic Change: A Toolkit for Professional Developers*. Like this *Toolkit*, it contains activities designed to be used by leaders, facilitators, or change agents to help people engage in conversations to create visions of effective mathematics and science instruction. (A description of this resource is included in the Resources section at the end of this chapter.)

The following summary of the literature on developing effective professional development comes from the sources mentioned above.

What Professional Developers Need to Know About Their Clients

A review of research conducted over the past 20 years reveals a consistent pattern of characteristics common to effective professional development. In reviewing over 200 studies on the subject of professional development, Showers et al. (1987) concluded that one of the most critical elements in planning successful professional development experiences is knowledge of the participants, including:

- Extent of content knowledge of the particular topic

- Skill level relating to topic
- Preferred teaching style(s)
- Preferred learning style(s)
- Readiness for change
- Perception of efficacy and ability to influence change, and
- Beliefs about students

An effective professional development opportunity will consider and address these characteristics whenever possible. *If the facilitator is unable to determine the above set of characteristics for the specific audience, initial activities that assess prior knowledge and expertise should be planned.*

Lessons from Adult Learning Theory

Lessons from the knowledge base on adult learning theory must also be considered in planning and designing effective professional development. What we know about how adults learn suggests that, at a minimum, attention must be given to the following items:

- Adults bring a history of previous learning experiences to any professional development event. These past experiences affect their attitudes and beliefs. *Professional development is most effective when it incorporates, recognizes, and validates these previous experiences.*
- Adults have varying styles of learning. *Professional development activities that incorporate a variety of learning modalities will be most effective.*
- Adults are motivated by practical applications and learning that is relevant to their own situations. *Professional development will be most effective when it is perceived to have practical use and direct application to the needs of those attending.*
- Adults are often busy, preoccupied, and required to address a variety of allegiances. *Effective professional development opportunities will be well-planned, efficiently organized, and present reasonable expectations and goals.*

Professional Development Models

A multitude of professional development design possibilities exist. In a review and synthesis of relevant literature and research, Sparks and Loucks-Horsley (1989) describe five models of professional development. Although their research focused on models of staff development for teachers, they can easily be adapted for administrators, parents, community members, or other stakeholders.

1. **Individually guided** models focus on the principle of "teacher-designed learning." In this model, a teacher identifies a need and through a variety of formal and informal procedures attempts to address this need. A middle school mathematics teacher might recognize the need to learn more about using discourse in his classroom. Through reading articles in professional journals, talking with other professionals, or locating and enrolling in a course on the topic, the teacher is involved in the individually guided model of professional development.

2. The **observation/assessment** model is characterized by feedback. This model is not limited to formalized evaluations, but includes observations and feedback by peers, supervisors, and others. While individual teachers can reflect on their own practice, research indicates this reflection is enhanced by outside observations. Another important research finding supporting the use of this model is the benefit incurred by both being observed and observing others. Using the observation/assessment model, a high school physics teacher attempting to teach students criteria for a lab report might ask a colleague to observe and provide feedback.

3. The model characterized by **involvement in a development/improvement process** suggests that teachers working with others will gain professional insights and skills through their participation in the process. A committee of elementary, middle and high school teachers assigned the task of developing mathematics performance assessments would exemplify this model. (The value, as professional development, of developing alternative assessments was discussed in Chapter 3, notably in **Activities 3.1** and **3.2**.)

4. The **training** model, although greatly overused, remains valid when used appropriately and integrated with a variety of other approaches. Workshops or institutes specifically designed to develop teachers' knowledge and skills provide valuable professional development opportunities. A workshop providing science teachers with improved knowledge in designing scoring criteria for alternative assessments typifies the training model. The "Short Term Training" section of this chapter provides other examples of this approach.

5. In the **inquiry** model, teachers working either individually or in groups become involved in active research. Through identifying questions, developing methodology, collecting data, discussing findings, formulating conclusions, and/or making recommendations, the inquiry model provides teachers with extensive opportunities for professional development. The "Vignettes" section of this chapter provides examples of this approach.

Specific Research on the Training Model

The extensive use of the training model has resulted in a large body of research. A review of this research and related literature suggests four essential components that must be considered and attended to in the planning and implementation of effective professional development training activities. Each component is critical and must be addressed in the design and delivery of this type of professional development opportunity.

1. **Theory.** An effective professional development training model provides participants with experiences that build understanding through a thoughtful examination of the theory supporting the experience. Professional development activities designed to address new assessment methods or practices must not only address the outward, visible signs of the strategy, but also the thought processes leading to the decision to select and use the strategy at the appropriate time.

2. **Demonstration.** Professional development training activities must provide participants with a concrete and coherent demonstration of the desired behavior. For example, if the purpose of the activity is to develop the ability to develop performance criteria, the facilitator must model how it is done and describe the rationale for each step.

3. **Practice.** Participants must have the opportunity to try out the skills and understanding they are developing through the professional development training activities. Participants must be given ample practice time during the activity to explore and build their understanding of the concepts and behaviors. Practice time is directly related to the successful adoption and adaptation of the concepts and skills by the participants upon their return to their local environment. However, practice must continue after the actual professional development event and be accompanied by constructive feedback.

4. **Feedback.** Feedback, both short and long term, is another critical component of effective professional development training models. As participants practice their new skills, both at the professional development event and back home, feedback is needed to ensure successful adoption of the target behaviors and knowledge. Facilitating participants' opportunities to practice and receive feedback after the initial professional development may be difficult, but is essential to success. Consistent research findings suggest that without follow-up practice and feedback, the initial successes of any professional development opportunity are severely diminished. Participants must recognize the need for continued practice and feedback, and be motivated to incorporate these opportunities into their working environment. Effective professional development training activities will either directly or indirectly facilitate feedback opportunities.

The Changing Paradigm of Professional Development

Just as perspectives of what constitutes good mathematics and science teaching are changing as a result of advances in cognitive research, professional development is undergoing similar changes. Barbara Spector (1993) compares characteristics of the old professional development paradigm to the new. She suggests, for example, that instead of focusing on correcting specific teacher behaviors, new programs should focus on teacher growth as a whole. Rather than asking teachers to passively sit and listen to a speaker, inservice programs should involve teachers in experiential, hands-on activities. And, rather than inviting university faculty to deliver speeches on some hot but perhaps fleeting topic, collaborative planning teams involving a wide variety of constituencies (e.g., higher education, elementary or secondary teachers and administrators, and other community leaders) should plan, design, and deliver sessions based on the specific needs of the group.

Further, the National Staff Development Council recently published *Standards for Staff Development* (1994), which notes that, contrary to earlier assumptions that professional development was only for teachers, current approaches "*recognize that everyone who affects student learning, from the board of education, central office administrators, principals, teachers, to classified support staff, and parents must continually improve their knowledge and skills in order to ensure student learning*" (p. 1). These standards, which are for middle school teachers (elementary and high school levels will follow), serve as guideposts for planning and conducting ongoing professional development activities in support of lifelong learning.

Table 7.1 lists characteristics of effective professional development that have repeatedly arisen in discussion sessions involving mixed groups of people concerned with improving mathematics and science education (e.g., classroom teachers, parents, administrators, members of the community, business representatives, state educators, and legislators). While some of the qualities listed have long been desirable attributes of professional development experiences, others have come to be expected (and demanded) more recently and reflect changes in approaches to professional development promoted by educational researchers. This list can be used to stimulate discussions among those planning professional development activities in mathematics and science education. It can also be used as a checklist to assess the effectiveness of a professional development experience both during (formative) and after (summative) the event. To apply it in this manner, a five-point Likert scale could be used to determine the extent to which a particular professional development event exhibited each characteristic (e.g., 1=not at all; 5=very much).

In sum, effective professional development training requires long-term contact, support, and feedback that results in empowered and reflective professionals who (according to Fullan, et al, 1990):

- Develop a *technical repertoire*—knowledge and skills that will contribute to successful implementation of assessment strategies

- Engage in *reflective practice*—thoughtful consideration of their assessment practices and how they influence other aspects of the educational enterprise
- Become *active researchers*—investigating, collecting data, analyzing and reporting results
- Seek *collaboration*—an activity that demands shared problem solving and a common vision

The long-term professional development vignettes at the end of this chapter serve as exemplars of successful professional development programs that have extended for two or more years and that demonstrate the above characteristics. All vignettes focus on mathematics and/or science alternative assessment.

Table 7.1

Characteristics of Effective Professional Development

Addresses attitudes, knowledge, skills	Integrates innovation
Aligns with school's strategic plan	Is interactive
Is applicable	Is interdisciplinary
Assessment/instruction are inseparable	Uses a variety of learning styles
Celebrates diversity	Models "less is more"
Challenges	Continues lifelong learning
Uses the change process	Is mediated by research findings
Establishes clear expectations	Involves metacognition
Is a collaborative effort	Models good teaching behaviors
Is comfortable	Moves from concrete to abstract
Encourages connection-making	Considers multiple intelligences
Uses constructivist ideas	Is needs-based
Is continuous	Is nonthreatening
Models cooperative learning	Includes performance assessment
Emphasizes depth over breadth	Is collaboratively planned
Is empowering	Exhibits professional value
Encourages creativity	Addresses real-life problems
Is engaging	Encourages reflection
Is equitable	Focuses on student learning
Is experience based	Maintains a supportive atmosphere
Is flexible	Allows for teacher growth
Uses follow-up support	Treats participants as professionals
Is hands-on	Takes a variety of forms
Is inquiry-based	Uses a variety of approaches

Sample Short-Term Training Plans

The activities contained in the *Toolkit* combined with the research findings reported in this chapter can be used as a foundation for planning professional development opportunities. The *Toolkit* activities can be "mixed and matched" depending on the identified needs of the participants and objectives of the particular professional development opportunity. Examples in this section cover a three-hour, a six- to seven-hour, and a two-day introduction to alternative assessment. Some assessment expertise on the part of the facilitator is helpful.

Three-Hour Introduction to Performance Assessment

The following sequence has been used successfully with teacher groups (elementary, middle school, and high school) as well as mixed groups of teachers, administrators and curriculum coordinators. (Note: this three-hour introduction specifically focuses on the more performance-based type of alternative assessment.)

Goals:

1. Become familiar with performance assessment terminology.
2. Become familiar with current performance assessment activities.
3. Begin to build a vision of quality with respect to performance assessment. Use this vision as a lens through which to view current alternative assessments.
4. Begin to build a vision of how performance assessment can be used in the classroom as a tool for student learning.
5. Understand the importance and classroom usefulness of good quality performance criteria.

***Toolkit* Resources:**

Chapter 2—Definitions; **Activity 2.5**—*Clapping Hands*; **Sample 4.2**—*Mathematics Problem Solving*; **Sample 4.4**—*Sow Bugs*; **Sample 5.1**—*Assessment and Technology Videotape*; **Sample 5.3**—*Science Portfolio (GSE)*; **Sample 5.9**—*Primary Math Portfolio*

Training Sequence:

1. Introductions, goals, agenda, participant questions.

2. Definitions—performance assessment, performance task, performance criteria. These three are the ones most needed in this workshop. (Use Chapter 2-*Definitions* as a handout.)

    a. After presenting the short definition of performance assessment ("assessment based on observation and judgment"), ask participants to think of examples both in daily life and in school. Point out that performance assessment is not new; what is new is the attempt to be more systematic so that performance assessment can be used for more things.

    b. Define performance task as "the activity given to the students to do." Ask participants to identify the task in such common situations as the driving test (parallel parking, backing around a corner, etc.), Olympic diving (three compulsive and three free dives), foreign language (speaking), etc.

    c. Ask participants to define "criteria." Summarize their comments with the definition in the handout. Ask participants to identify the criteria for the situations in (b) above. Also have them think of criteria for restaurants, hotels and other familiar situations. Ask them: "Would it be important to know the criteria for the driving test...ahead of time?" Likewise, would it be useful for students to know ahead of time the criteria for success in school-related areas, such as writing, problem solving, and group collaboration? Why? (If we don't know the criteria for success, how can we be successful?)

3. Try out one or two performance assessments. Give or show the participants a student performance (for example, **Sample 4.2**—*Mathematics Problem Solving*). Ask them to rate the quality of the performance without criteria. Then show them the criteria and ask participants to score the performance again. This leads to discussions of the necessity for performance criteria. Also, have participants identify the task and the criteria and discuss whether it would be useful for students to have the criteria in advance.

4. Point out that there is no single right way to do performance assessment, just as there is no single right way to teach. However, our decisions on how we do performance assessment will have consequences. Do **Activity 2.5**—*Clapping Hands* to illustrate what can happen when we design performance assessment in various ways. This leads to a discussion of quality in performance assessment.

5. Use the discussion of quality as a lens through which to review and critique sample performance assessments. Pick three to six samples from Appendix A, preferably a variety of on-demand assessments, projects, and portfolios. For example, use **Sample 4.2**—*Mathematics Problem Solving* or **Sample 4.4**—*Sow Bugs* to illustrate on-demand assessments; use **Sample 5.1**—*Assessment and Technology Videotape* to illustrate a project; and use **Sample 5.3**—*Science*

Portfolio (GSE) or **Sample 5.9**—*Primary Math Portfolio* to illustrate portfolios.

6. End the session by asking the participants to self-reflect using three questions: "What did I learn?" "What am I still confused about?" and "What am I going to try?"

Six-to-Seven Hour Training Event on Performance Assessment

The first half of the session is identical to the above. The following portions are added:

Additional Goal:

6. Practice developing performance criteria

Additional *Toolkit* Resources:

Activity 2.9—*Going to School*; **Activity 3.1**—*Camping Trip*; and **Activity 4.3**—*Performance Criteria—Keys to Success*

Additional Training Steps:

6. Review the necessity of having clear criteria both for assessment and for use with students in the classroom to communicate what it takes to be successful. **Activity 2.9**—*Going to School* can be used.

7. Use **Activity 3.1**—*Camping Trip* to show participants how to develop performance criteria.

8. Use **Activity 4.3**—*Performance Criteria—Keys to Success* to summarize and deepen participant understanding of the characteristics of good performance criteria.

9. End the session by asking the participants to self-reflect using three questions: "What did I learn?" "What am I still confused about?" and "What am I going to try?"

Two-Day Training Sequence on Performance Assessment

The two-day event builds on the one-day event.

Additional Goals:

7. Explore the characteristics of sound performance tasks

8. Plan a performance assessment to meet your own needs

Additional *Toolkit* Resources:

Activity 4.1—*Performance Tasks—Keys to Success* and Chapter 2 (student outcomes best assessed using performance assessment)

Additional Training Steps:

9. Participants now have enough background that they can begin to plan their own assessments. The first two steps they need to work through are clarity of purpose of assessment and deciding what they want to assess. Performance assessment is not appropriate for all our goals for students. Use the information in Chapter 2 to help participants choose a student learning goal that is appropriate to assess using performance assessment.

10. Once participants have chosen a student skill to assess, help them find performance criteria that can be used to assess performance on that skill. For example, if the focus is math problem solving, show them samples of performance criteria others have used to assess math problem solving. If they can't find criteria, they will need to plan how they will develop criteria. (Refer back to **Activity 3.1**—*Camping Trip.*)

11. The final design consideration in a performance assessment is tasks. Use **Activity 4.1**—*Performance Tasks—Keys to Success* to begin to think about the characteristics of quality performance tasks. Then have participants find, adapt or develop a performance task to elicit the skill they want to assess.

12. End the session by asking the participants to self-reflect using three questions: "What did I learn?" "What am I still confused about?" and "What am I going to try?"

Long-Term Professional Development Vignettes

School Community-Based Management (SCBM) Projects

In the state of Hawaii, the Pacific Mathematics and Science Regional Consortium is supporting and documenting an emerging model for teacher professional development. The model combines professional development, staff credit courses, cognitive coaching, focused teacher task forces, action research by teachers, and an "alignment center" component.

The implementation of this model is facilitated by Pacific Consortium staff and partners from the University of Hawaii working in residence at the school site one to two days per month to provide focused technical assistance to small groups of teachers. The technical assistance addresses a variety of needs: (1) developing thematic curriculum units, (2) creating active learning environments, (3) using heterogeneous work groups, and (4) developing quality classroom assessments. On the "resident" days, teacher groups meet by grade level during "dialog time" made available by the school's decision to eliminate specialist positions in conjunction with a goal to enrich learning for all students. In place of "pull-out" programs, part-time instructors work with all students, allowing blocks of time when the classroom teachers can work and plan together.

The assessment work currently under way includes (1) developing a portfolio framework based on a school-community vision for all students' success, (2) creating and using student generated criteria for effective group work, and (3) examining and creating alternatives to the current reporting system. Areas of special interest include concept mapping as an assessment tool, the use of extended observation checklists, self-reflection logs, and teacher journals. This work is the result of a vision jointly created by teachers, students, parents, community members, school support staff, and school administrators.

Yap Classroom Learning Assessment Process

This model for professional development includes long-term, sustained support from within the school system for development of assessment skills and knowledge. It is intended to respond to the need for professional development across great distances and is suited for use in rural and isolated environments. Designed by local educators to help teachers develop and use quality assessment tools to portray the learning that is valued by the people of Yap State, the model involves systemic reform across curriculum, instruction, and assessment.

Yap State in the Federated States of Micronesia includes schools located on multiple islands in the western Pacific. A core group of five local trainers/coaches takes part in multi-phased, intensive training of trainers focused on both assessment content and coaching strategies to facilitate learning. The core group is made up of central staff, including the curriculum and instruction supervisor, mathematics and science specialists, a specialist in a community school-based improvement process, and members of the professional development staff.

Each phase involves gradually moving training and coaching responsibilities from external resource people (Pacific Regional Educational Laboratory [PREL] staff) to Yap's core group. Phase I built on a vision for Yap's students, connected to Yap's science and mathematics curriculum frameworks, and focused assessment activities on guidelines for paper and pencil assessment, assessing student thinking, and introducing performance assessment. The first step was a model workshop for trainers and teachers facilitated by PREL staff and based on needs and priorities set within Yap State. Moving to Ulithi atoll, the local trainers selected activities and used their knowledge of culture and curriculum to adapt the design, co-training with PREL staff. Gaining experience and confidence, Yap's trainers conducted the training with additional schools in a third neighboring island location (Woleai atoll).

Phase II was called "Curriculum Connections" and began with an examination of assessment questions raised during Phase I in relation to the curriculum and effective teaching. Phase III explored classroom questioning—opportunities for using personal communication for assessment. During this phase, the Yap trainers incorporated coaching strategies with teachers. Yap's trainers have designed and implemented school-based workshops capturing the essential content of each phase and have been conducting sessions throughout the summer months.

Similar multiphase assessment projects have begun in several Pacific entities. Teachers are the trainer/coaches in Pohnpei state. All have built-in independent action following intensive training of trainers. The model evolves as trainers and teachers grow and make the content of effective assessment their own.

Applying Alternative Assessment to District-Developed Science Kits: A Case Study

Background

In the summer of 1991, a mid-sized Texas school district engaged teachers in a project to develop science activity kits using the community's available facilities, resources, and geographic features. With the support of the superintendent of instruction, local and Eisenhower funds were secured to hire a district science specialist to direct the project. The goals of the resulting three-week summer workshop and year-long follow-up sessions were to provide professional development experiences in science for teachers and to have them apply their new knowledge directly in writing appropriate activities for the science kits. The kits, which were based on the Texas Essential Elements (the state learner objectives), followed models of other kit projects from throughout the country in that each activity was accompanied by an assessment, usually traditional in form.

During field testing of the kits, the teachers experienced some frustration—although the activities they had designed were hands-on, the assessments did not follow the same format and did not allow the students to demonstrate what they had actually learned. However, none of the teachers had any formal training in designing assessment items, so they lacked the confidence to produce appropriate tasks, criteria, and rubrics.

At the same time, the state of Texas was considering revising its statewide science assessment to be more performance-based. In fact, districts were informed that there was a strong likelihood that at least one performance task would be included on the 1993-94 Texas Academic Achievement Skills (TAAS) test. As it turned out, a TAAS performance task was developed, but was optional and ultimately designated for district use only. Nevertheless, the leanings of the state provided an added impetus to revise the science kit assessments. In the spring of 1993, the district competed for and was awarded a three-year professional development mini-grant from the Southwest Educational Development Laboratory (SEDL) to simultaneously train teachers in the latest alternative assessment procedures and have them apply their knowledge to revising the science kit assessments.

Project Design

The district science specialist, Tom, and a university education professor, Michael, attended an initial week-long training-of-trainers session conducted by SEDL for mini-grant recipients, in which current research topics in mathematics and science education (e.g., constructivism, integration of mathematics and science, cooperative learning, less is more, peer coaching, reflection time for processing information, and observation and journal writing as assessment approaches) were explored, modeled, and practiced. Mini-grant recipients subsequently designed summer workshops for teachers, incorporating the suggested approaches as appropriate.

In designing their summer workshop, Tom and Michael had to work within the constraints of university requirements. Because the group met minimum size requirements, Michael was able to secure approval to hold formal classes at the district's science resource center rather than on the university campus. Further, he worked with Tom to custom design a course to meet the needs of the teachers, the district, and SEDL. The 18 elementary science teachers recruited for the project had the option of receiving university credit for their participation. To receive the credit, they were required to log a minimum number of contact hours with Michael, attend sessions, do assigned reading, keep journals and portfolios, conduct and videotape performance assignments in their classrooms, and report back to peers on the results. Only a few of the participants had actually been involved in the previous kit-development project.

Tom explained the purpose of the science kit alternative assessment training project as follows:

> *The content of this training focuses on developing conceptual understanding of assessment strategies appropriate to five domains of science education. . . The operating principle of our training approach is "experience it first.'" As such, teachers in the training sessions encounter new assessment tools as users first. This experience base yields opportunities to discuss the tools' salient features and to apply tools in a classroom setting.* (Tom)

Project Activities

Project activities have included the following:

- After exploring either sample video-disc-based assessment instruments or selected activities from the New York State Department of Education's Elementary Science Manipulative Skills Test, teachers administered one of the enriched tasks to 8-10 students and reported on the results.

- After meeting with a Project 2061 consultant to discuss constructivism, back mapping, benchmarks, and rubrics, teachers developed and administered a performance task (with rubrics) to determine the extent to which a selected benchmark had been met.

- After pilot testing selected activities from the *Toolkit* (e.g., **Activity 2.2**—*Pictionary* and **Activity 3.1**—*Camping Trip*), teachers explored the rationale for developing scoring criteria and rubrics, developed a performance task to allow students to demonstrate learning, had students complete the task, and collected student responses, which were subsequently incorporated into a new *Toolkit* activity (**Activity 3.2** – *Sorting Fish*)

- In pairs, teachers selected one or more district science kits, created alternative assessments for the kit(s), pilot tested the assessments with their own students, reported results of their pilot tests, and modified the assessments as necessary.

For each assignment, teachers provided portfolio samples of student assessment work, samples of student responses, teacher observations of the assessment, an estimation of class time allotted to the assessment event, and an overview of efforts to link the assessment event with ongoing instruction. Teachers also shared their experiences in small group settings.

In one such sharing session, teachers directly addressed many of the problems of the practical application of an assessment approach. Following a *Toolkit* activity, they developed criteria for a science task, rubrics, and the task itself. They then administered the task to their students, and brought the student papers to the training session. After exchanging papers (without rubrics), teachers in small groups categorized them according to levels of student response and determined a rubric based on those responses. They compared this rubric with the teacher-developed rubric. The ensuing discussion focused on pros and cons of sharing a rubric with students before administering the task and of using a three-, five-, or seven-point scale. Throughout these discussions, the issues of validity and reliability surfaced repeatedly.

In another training session, teachers worked in pairs or small groups to design and pilot test assessments for the district science kits. This process was not always straightforward:

> *We decided some of the activities in the kit were poor and had problems...[and] that if the rubrics we made were linked to poor activities, the rubrics were no good either. We worked on changing and adding to the weak activities...then altered some of our rubrics.* (Beth)

Project Impact

The teachers first learned about alternative assessment in science and then developed and field tested alternative assessments for the district's science kits. The ongoing project thus focuses on alternative assessment as a vehicle for improving science instruction. Both teacher training sessions and follow-up applications of assessments in the classroom were monitored by a SEDL liaison to assess the effectiveness of the professional development project. The following elements have contributed to its success:

1. The collaborative nature of the project:

- School principals have allowed a minimum of two teachers per school to participate, thus encouraging peer support.

- The district has served as fiscal agent for the project and provided direction, planning, trainers, substitute teachers, meeting sites, services, and the resources of the project director and Science Resource Center.

- The university has provided higher education credit for the teachers, in addition to a professor as a trainer who was able to tailor the course for "student" needs and deliver it off campus.

- SEDL has provided training of trainers, technical assistance, and funding for teacher stipends, substitutes, consultants, tuition, materials, and travel.

2. The opportunity and time to apply immediately the theoretical and practical knowledge gained—testing prepared performance assessments, and developing and testing alternative assessments for the district's science kits

3. The recruiting of pairs of teachers from individual schools to further on-campus collegiality

4. The use of a Project 2061 consultant to assist teachers in linking science benchmarks and assessment

5. Field trips—to the Psychological Corporation to gain firsthand experience in how standardized tests are developed and tested and to a nearby school district successfully implementing portfolio assessment—and an opportunity to share perceptions with colleagues following the trips

6. The long-term nature of the project, with initial training, application, peer interchange, reflection, and follow-up support (the project director has regularly visited the classrooms of participating teachers, offering guidance and feedback on request)

Although teacher participants earned university credit in May 1994, the project continues. At their schools the teachers have planned and presented awareness sessions to fellow teachers, administrators, and parents on effective use of the alternative assessments in the science kits. They have also been further revising the science assessments using teacher feedback.

In the fall of 1994 the teachers were polled for their preferences regarding the focus of activities for the final project year. In addition to continuing their refinements of alternative assessments for the science kits and acting as mentors for teachers from schools that have not participated in the alternative assessment project, participants (perhaps because of their exposure to the video-disc-based science assessment) wanted to learn more about integrating science and technology. Original training plans have been modified to accommodate these preferences.

Six months into the project, teachers were asked how their participation had changed what they do in the classroom:

> *I am more aware of different techniques used in testing. . . [and] more sensitive to the different learning styles in my classroom.* (Sandy)

> *I always state my grading criteria when an assessment begins. I am more project-oriented now.... I really feel like working with these alternative assessments has broadened my range of success with the students.* (Paula)

A year into the project, participants reported:

> *When we began we were fairly confident in our use of "hands-on" activities. Very quickly we realized our knowledge did not convey into the realm of assessing those same activities.* (Cheryl)

> *Participation in this grant has allowed me to become a stronger science teacher. I have always felt like science should be more hands-on but never really knew what to do to justify the "grading" of such hands-on activities. Now I have the tools to do this.... It is wonderful to be able to assess student knowledge in a way other than the traditional assigning of a number.* (Michelle)

> *Since I have worked with these assessment tools, I feel better qualified to teach science. I like the idea that the students know up front what the objectives are and what they are expected to know. I feel the students are now better prepared to face the real world because they are expected to be actively involved in science using hands-on. My classroom is very busy and active now. It's opened up a new world.* (Jane)

> *Prior to joining this group, I. . . believe[d] in hands-on science. . . [but] was struggling with ways to get a grade. . . The district demands a justified numerical grade. Thanks to. . . [this project], I now have several superior ways to grade my students. Not only that, but I can tell exactly what my students gained out of every lesson.* (Melissa)

> *A whole new form of assessment was introduced to me. I remember when we did practice assessments on ourselves and I kept asking myself, how can we measure curiosity? Now I know! In my eyes, using these assessments (rubrics, enriched multiple choice, performance tasks). . . [is] a way of grading that makes sense. It has changed my way of thinking and my way of teaching. I find myself looking for other forms of grading my students. My students have a chance in science (and other subjects) to not only write down what they know but they can also show me what they know. . . It has taken me some time to change my mind set.* (Sandy)

> *I've learned in a greater depth my content area of science during the process of creating rubrics by focusing on what the children need to be able to conceptualize, not just memorize. I've also seen that using rubrics creates a*

> *positive environment where students can grow successfully because they feel confident in what is expected of them. This makes learning and teaching better for all. It makes much more sense than playing the guessing game as to what do you think will be on my test?* (Beth)

Perhaps the best indicator of personal change can be gleaned from the videotaped sequences and journal entries of a veteran fourth-grade teacher, Allen. In initial journal entries, he was skeptical about the applicability of alternative assessment to his classroom. Subsequent entries show a progression from an attitude of "*Okay, they might have some use,"* to *"I wonder what parts of this are applicable to me,"* to *"It works and what else do I need to do to make it work better?"* His most recent entry reads:

> *My personal growth and development resulting from affiliation with this grant has become most noticeable to me as I read through lesson plans and activities that I had developed prior to this training. In my older plans, I used some hands-on, some laser disc, movies, video tapes, and text book exercises; but I tested my students in traditional ways, either true or false, fill-in-the-blank, etc. I realize now I had probably done my students an injustice. I now use the system we have learned, using rubrics, grading on performance, and using a more simple format. I find that my students now feel less pressure, are responding to instruction in a more positive manner, and I have significantly less. . . failures. . . I have become a much more observant teacher and facilitator. I. . . also feel less pressured because I evaluate performance as the students are following instruction and am not spending many after-school hours grading test papers.*

The experiences of these teachers illustrate the gradual changes that must occur for teachers to embrace the new assessment paradigm. Such changes require the support of both administrators and colleagues, and can be greatly enhanced through collaborative efforts involving school districts, institutions of higher education, and other entities from both within and outside of the community.

References

Progress Report & Continuation Proposal, August 1993.

SEDL site visit notes, September 27-28, 1993, and April 18-19, 1994.

Teacher journal entries, February 1, September 8, and September 13, 1994.

Telephone interview, Tom, November 3, 1994.

Videotaped classroom assessment scenes, September 1993 and May 1994.

Additional Reading and Resources

This alternative assessment *Toolkit* is just one resource to be used in the development and implementation of effective professional development.

Another major resource is the Regional Educational Laboratory Network Program product, *Facilitating Systemic Change: A Toolkit for Professional Developers* (in development by The Regional Laboratory for Educational Improvement of the Northeast and Islands). Following a national search for programs and practices that have the potential of transforming mathematics and science learning, *Facilitating Systemic Change* was designed to help create a link between promising programs and practices and educators trying to translate them into classroom practice. It is a cutting-edge resource for science and mathematics education leaders who are serious about furthering reform agendas in five areas: Setting the Stage for Reform; Effective Science and Mathematics Education; Effective Dissemination Strategies; Effective Professional Development; and Effective Change Strategies.

Two critical, interrelated themes are woven throughout the *Toolkit:* using a systemic approach to change and ensuring equitable access to science and mathematics for all students. The first theme stems from the realization that a new kind of thinking is needed if the current reform is to succeed where past reforms have failed; thinking that considers all parts of the system rather than one at a time; pays as much attention to the ways the parts are linked as to the parts themselves; and focuses on efforts to redesign the system based on a clear, shared vision of what it takes for all learners to succeed. The professional development *Toolkit* contains 18 activities (including relevant research, training agendas, overheads, and facilitator guidelines) to conduct professional development events.

For more information on *Facilitating Systemic Change: A Toolkit for Professional Developers*, contact your Regional Educational Laboratory Network Program representative.

Other references include:

Fullan, M., Bennett, B., & Rolheiser-Bennett, C. (1990). Linking classroom and school improvement. *Educational Leadership, 47*(8), pp. 13-19.

Garmston, R. & Wellman, B. (1992). *How to make presentations that teach and transform.* Alexandria, VA: ASCD.

Kober, N. (1991). *EdTalk. What we know about mathematics teaching and learning.* Washington, DC: Council for Educational Development and Research.

Kober, N. (1993). *EdTalk. What we know about science teaching and learning.* Washington, DC: Council for Educational Development and Research.

Little, J. (1986). Seductive images and organizational realities in professional development. In A. Lieberman (Ed.), *Rethinking school improvement.* New York: Teachers College Press.

Loucks-Horsley, S., et al. (1987). Critical attributes of effective teacher development programs. In *Continuing to learn: A guidebook for teacher development.* Andover, MA: Regional Laboratory for Educational Improvement of the Northeast and Islands.

McLaughlin, M. & Marsh, D. (1980). Staff development and school change. *Teacher College Record, 80*(1), pp. 69-93.

National staff development council's standards for staff development. (1994). Oxford, OH: National Staff Development Council.

Olson, T., Butler, J., & Olson, N. (1991). *Designing meaningful professional development: A planning tool.* Portland, OR: Northwest Regional Educational Laboratory.

Regional Educational Laboratory Network Program. (1994). *Improving science and mathematics education—A toolkit for professional developers: Alternative assessment.* Portland, OR: Northwest Regional Educational Laboratory.

Showers, B., Joyce, B., & Bennett, B. (1987). Synthesis of research on staff development: A framework for future study and a state-of-the-art analysis." *Educational Leadership, 45*(3), pp. 77-87.

Sparks, D. & Loucks-Horsley, S. (1989). Five models of staff development for teachers. *Journal of Staff Development, 10*, pp. 40-55.

Spector, B. (1993). Order out of chaos: Restructuring schooling to reflect society's paradigm shift. *School Science and Mathematics, 93*(1), pp. 9-19.

Stake, R. (1987). An evolutionary view of programming staff development. In Wideens & Andrews (Eds.), *Staff development for school improvement: A focus on the teacher.* Philadelphia, PA: The Falmer Press, Taylor and Francis, Inc.

Yarger, S., Howey, K., & Joyce, B. (1980). *Inservice teacher education.* Palo Alto, CA: Booksend Laboratory.

Workshop Evaluation Forms

Two workshop evaluation forms are included: *Documentation and Evaluation Form* and *Workshop Evaluation Form*. The former is intended for use by professional development coordinators or trainers to self-reflect on their training events and on *Toolkit* activities. The latter is for use by workshop participants to provide feedback to coordinators and trainers.

A Toolkit for Professional Developers: Alternative Assessment

Workshop Presenter

◆ Documentation and Evaluation Form ◆

Please complete this form **after** you use material from the *Toolkit*. (If you send feedback to the *Toolkit* developers, we can use your comments to modify the next edition.) Thank you.

1. Facilitator(s): ____________________ 2. Position/Title ____________________

3. Organization ______________________________ 4. Date ________________

5. What was the **title** of the overall event in which the material was used? ________________

____

6. What was the **overall purpose** of the sessions? ______________________________

____

7. Please state the **professions of participants**, (e.g., principals, curriculum directors, etc.) and **numbers** of each:

____

8. What keynotes, workshops, etc. took place **in addition to** the *Toolkit* activities?

____

9. Materials used (chapter, activity number[s], and/or samples): ________________________

____

10. Why did you select this particular material from the *Toolkit*? ________________________

____

____

11. What was most valuable about each item used? ______________________________

____

____

(next page)

12. What was most problematic about each item used? ______________________

____

____

____

13. How could the material be improved? Please attach any additional or substitute handouts, transparencies, processes, readings, etc. that you used or would use next time.

____

____

____

____

14. What follow-up, if any, will there be to this use of the *Toolkit*? ______________________

____

____

15. What issues, concerns, or ideas would you like to share with *Toolkit* developers or other users as a result of this experience? ______________________

____

____

____

16. Have you used the materials with other groups? ☐ yes ☐ no

17. Have you used other materials from the *Toolkit*? ☐ yes ☐ no

18. If yes, what have you used? ______________________

____

Please return to your Laboratory Network Program participant listed at the front of the *Toolkit* or to Judy Arter, NWREL, 101 SW Main St., Suite 500, Portland, Oregon 97204.

THANK YOU

Science and Mathematics Alternative Assessment: Workshop Evaluation Form

1. Please rate the quality and/or usefulness of the following aspects of the workshop:

	Very Useful/ Good Quality				Not Useful At All/ Poor Quality
a. Materials	5	4	3	2	1
b. Activities	5	4	3	2	1
c. Information	5	4	3	2	1
d. Presenter(s)	5	4	3	2	1
e. Pacing	5	4	3	2	1
f. Other: ________________	5	4	3	2	1

2. How do you plan to use the information from the workshop?

3. What was the best part of the workshop?

4. What suggestions do you have for improving materials and activities?

5. Other comments or ideas?

Please return this completed form to the workshop presenter before leaving today or mail to:

[your address label here]

THANK YOU!

Appendix A

Alternative Assessment Sampler

Index

About the Sampler

For convenience, all sample assessments have been placed together in Appendix A. Sample numbers refer to the primary chapter in which they are used. For example, Sample 4.5 is especially relevant to Chapter 4. In order to assist you to choose the sample assessments you might want to use in various activities, the following information is provided about each sample. (Note: our descriptions apply to the particular sample included. Other parts of the same assessment might have other characteristics.)

1. **Title, source.**

2. **Grade level, subject area emphasized, level** (classroom or large-scale assessment).

3. **Task description.** Group and/or individual work; type of activities (written or hands-on); type of student responses (written, oral, visual, or physical); number of right answers (a measure of degree of open-endedness); complexity (short answer/constructed response, on-demand performance assessment, project, or portfolio); whether self-reflection is included.

4. **Description of performance criteria.** General or task-specific; holistic (one-score) or analytical trait (more than one score); what is covered by the criteria; the amount of detail and description.

5. **Student work.** Whether we have included sample student work from the assessment. (Other sample student work may be available from the authors.)

Samples

Sample 3.1 **Acids-Bases Assessment Task—Cleaning Up a Toxic Spill**
Far West Eisenhower Regional Consortium

Science, grades 7-12, classroom assessment. Tasks—hands-on; written-response; project; individual and group; self-reflection. Performance Criteria—task-specific; five scores for different parts of the experiment. Samples of student work included.

Sample 4.1 **How Many Buttons?**—*Mathematical Sciences Education Board*

Math, grade 4, classroom assessment. Task—hands-on; individual and group; project; more than one right answer. Performance Criteria—task-specific; one score based on reasonableness of approach, communication, and correct computation; descriptive. Samples of student work included.

Sample 4.2 Mathematics Problem Solving—*Oregon Department of Education*

Math, grades 3, 5, 8, and 11, large-scale assessment. Tasks—more than one right answer; written problems and responses; individual; on-demand. Performance Criteria—generalized; four scores—understanding the problem, procedural knowledge, problem solving, and communication; descriptive.

Sample 4.3 Discovering the Problem of Solid Waste—*Lake County ESC, Illinois*

Science, grade 4, classroom assessment. Task—application of knowledge; several questions organized around a theme; short answer; both one right answer and more than one right answer; hands-on; written responses; individual work. Performance Criteria—task-specific; one score for each question, based on correctness or reasonableness of response; not descriptive.

Sample 4.4 Sow Bugs—*Richard Shavelson, University of California at Santa Barbara*

Science, grades 7-9, classroom assessment. Task—one right answer; hands-on; individual; on-demand; written response. Performance Criteria—task-specific; several scores; performances observed for science process skills.

Sample 4.5 Weathercaster's Helper—*Utah Department of Education*

Science, grade 1, large-scale assessment. Tasks—one right answer; questions centered around a theme; short answer; written problems and responses; individual. Performance Criteria—task specific; one score for each answer; question scores totaled to form subtest scores for science process skills.

Sample 4.6 Aquarium Problem—*New Standards Project*

Math, grades 4-6, level of assessment not specified. Tasks—more than one right answer; written problem and responses; individual; on-demand. Performance criteria—task-specific; one score based on reasoning, understanding, and accurate use of information; descriptive. Samples of student work included.

Sample 4.7 Performance Assessment in Math—*Alberta Education, Canada*

Math, grade 9, large-scale assessment. Task—one right answer; hands-on; written response; individual; on-demand. Performance criteria—generalized; two scores—problem solving and communication; descriptive.

Sample 4.8 Science—New Directions in Assessment—*California Department of Education*

Science, grades 5, 8, and 10, large-scale assessment. Task—more than one right answer; hands-on; written responses; individual; on-demand; several activities around a common theme. Performance criteria—generalized; descriptive; one score based on science process skills such as valid observations and good data displays.

Sample 4.9 Mapping the Blue Part—*Kentucky Department of Education*

Science, grade 8, large scale assessment. Tasks—several questions on a common theme; some have one right answer and some have more than one right answer; hands-on; written responses; both individual and group work; on-demand. Performance criteria—generalized criteria that are tailored to individual questions; some description; one score for each question based on conceptual understanding, efficiency, sophistication, and insightful interpretations. Samples of student work included.

Sample 4.10 Writing in Chemistry—*CRESST*

Science, high school, level of assessment not specified. Task—more than one right answer; written problem and responses; individual; on-demand. Performance criteria—generalized; some description; four scores—use of concepts, quality of argumentation, use of information from sources provided, and number of misconceptions; some scores based on counting frequency of features.

Sample 5.1 Assessment and Technology Videotape—*Center for Technology in Education, Bank St. College of Education*

Physics, grade 11, classroom assessment. Tasks—hands-on; oral and written responses; more than one right answer; group work; project; self-reflection. Performance Criteria—general; seven scores: understanding, critical thinking, oral presentation, written work, teamwork, effort, reflectiveness; some description.

Sample 5.2 Assessment of Laboratory Skills in High School Science
Rodney Doran, State University of New York at Buffalo

Chemistry, grade 12, classroom assessment. Tasks—on-demand; hands-on; individual work; written response; range of best answers. Performance Criteria—general; 10 scores covering experimental design and reporting experimental results; some description.

Sample 5.3 **Science Portfolio (GSE)**—*California Department of Education*

Science, grade 12, large-scale assessment. Tasks—portfolio; more than one correct answer; self-reflection; written, visual, oral responses. Performance Criteria—general; one score for each of three portfolio entries; scores based on insight, reflection, conceptual mastery, reasoning, and communication; descriptive. Samples of student work included.

Sample 5.4 **Grade 3 Interdisciplinary Task**—*Maryland Department of Education*

Interdisciplinary (reading, science, writing), grade 3, large-scale assessment. Tasks—several short answers on a common theme; group and individual; hands-on; some questions have more than one right answer. Performance Criteria—task specific or general depending on the question; scores based on observation, communication, understanding, critical stance, language use, and expository writing.

Sample 5.5 **Assessment of Learning and Communication Processes**
Alberta Education, Canada

Cross-disciplinary (science, social studies, language arts), grades 7-10, classroom assessment. Tasks—project; more than one right answer; individual and group; hands-on; several response modes. Performance Criteria—general; six scores for different learning and study skills; some description.

Sample 5.6 **Algebra II**—*Michael Lehman, Holt Public Schools*

Algebra, grade 10, classroom assessment. Task—on-demand; group; written problems; oral responses; group work; more than one right answer. Performance Criteria—general; 10 scores on quality of the math, problem solving, communication skills, and group process skills; sketchy description.

Sample 5.7 **Mathematics Assessment**—*California Department of Education*

Mathematics, grades 4, 8, 10, large-scale assessment. Task—written problems; written responses; individual; more than one right answer; on-demand. Performance Criteria—General; one score based on conceptual understanding, problem solving, and communication; descriptive. Samples of student work included.

Sample 5.8 **A Day at the Carnival**—*Utah Department of Education*

Mathematics, grade 6, large-scale assessment. Task—individual; several questions on a common theme; short answer; mostly one right answer; written problems and responses. Performance Criteria—task-specific; one score for each question; question scores totaled to form subtest scores for science process skills.

Sample 5.9 Primary Math Portfolio—*Leon Paulson, Multnomah Educational Service District*

Mathematics, grade 1, classroom assessment. Task—portfolio; individual; self-reflection. Performance Criteria—General; one score for the overall sophistication of the portfolio based on coherence, purposeful, complete, awareness of audience; some description.

Sample 6.1 Hawaii Algebra Grading Process—*University of Hawaii*

Algebra, grade 8, classroom assessment. Task—individual journal writing and group problem solving; written and oral response; more than one right answer. Performance Criteria—journal entry gets one score for completeness, communication, and quality; group problem solving gets 16 scores for ideas, organization, presentation skills, and group process skills.

Sample 3.1

Acids-Bases Assessment Task— Cleaning Up a Toxic Spill

Far West Eisenhower Regional Consortium

The following assessment task is embedded in a unit in which students work together and experiment to learn about neutralizing acids and bases. A five-day sequence of hands-on lessons begins with students mixing common household items with a pH indicator and recording the results. The second day, students discuss their previous experiments and participate in a demonstration of the indicator color spectrum using a flask, indicator, ammonia, and vinegar. Using household chemicals brought in by the students, the teacher reinforces the concept of pH spectrum and introduces the experimental problem of comparing acid strengths of vinegar and aspirin. On day three, after discovering that both aspirin and vinegar turn the indicator a pink/red color, students experiment to learn which of the two is a stronger "pink-maker." They also create a spectrum of colors using drops of vinegar and ammonia in test tubes containing indicator, reinforcing that acids and bases neutralize each other. The next day students connnect concepts of acids and bases to real world examples such as acid rain, tooth decay, cabbage juice indicator, and pH in our bodies.

The assessment task "Cleaning Up a Toxic Spill" is given on the fifth day, when students work with a partner to design and conduct an experiment to determine which of four "mystery" cleaners is the strongest base. Prior to engaging in the assessment task, students discuss the criteria on which their work will be judged—how well they design an appropriate experiment; carry out the experiment; describe what they are going to do and what they did; draw the correct conclusions from their data; and explain their conclusions. Students become familiar with a self-scoring sheet which they will use after the assessment to describe how well they think they did on each part of the task.

The following pages contain the assessment task, the task-specific scoring rubric, student self-scoring sheet, and sample student work from three sets of partners.

From "Altering Assessment Workshop," 1994
Far West Eisenhower Regional Consortium
730 Harrison Street
San Francisco, California 94107

Permission for use in training has been granted by Far West Eisenhower Regional Consortium. Materials may be copied for training purposes only.

STUDENT NAMES ____

Cleaning Up a Toxic Spill

You are an Environmental Scientist. A truck has an accident on the highway and a barrel of a chemical called Big Messium spills on the road. The Highway Police ask for your help. You look up Big Messium in your computer and learn that the best way to clean it up is to use a strong base. You have four different cleaners that you are allowed to use to clean up highways. You need to do an experiment to decide which is the strongest base.

Supplies

Safety glasses – wear these when handling or near the solutions

Masking tape

Ten test tubes in test tube holder

Four labeled Cleaners in dropper bottles

Dropper bottles with **pH Indicator** and **Vinegar**

1) First you need to find out if any of the cleaners are a base. Describe in one or two sentences the **simplest** experiment that you can do to find out which of the cleaners are bases.

2) Do the experiment and record your results in the Data Table.

Cleaner	Color Observations	Is It A Base?

3) If more than one cleaner is a base, you need to do an experiment to discover which is the strongest base. **What materials are you going to use? Write step-by-step what you are going to do. Use a picture if that helps.**

4) Do your experiment. Make a Data Table to show your results.

5) Which of the cleaners is the strongest base?

6) What evidence proves that it is the strongest base?

Permission for use in training has been granted by Far West Eisenhower Regional Consortium. Materials may be copied for training purposes only.

CLEANING UP A TOXIC SPILL — PARTNER TASK — SCORING RUBRIC

The level of performance centers around the concept of "there" which refers to the respondent(s)' demonstrating where we want them to be with respect to the dimensions of performance associated with this task in terms of experimental design, experimental procedure, analysis of results/drawing conclusions, and argumentation/support of conclusions.

This rubric is to be used for any item to which there is a response. *Score all unanswered items as 0* (no response).

DIMENSION OF PERFORMANCE	LEVEL OF PERFORMANCE		
	3 THERE	2 GETTING THERE	1 NOT EVEN CLOSE
I. DETERMINATION OF BASES	• correctly identified the **3 bases and the 1 non-base** ***by combining cleaner & indicator*** • ***did not*** use vinegar	• correctly identified the **3 bases and the 1 non-base** ***by combining cleaner & indicator*** • ***did*** use vinegar	• did not correctly identify the **3 bases and the 1 non-base** • **response** indicates procedure other than those in Levels 2 & 3
II. DESCRIPTION OF EXPERIMENT	• procedure used is ***fully clear*** from either the text (with or without graphics) or the data table or a combination of the two	• procedure used is ***partly clear*** from either the text (with or without graphics) or the data table or a combination of the two	• procedure used is *unclear* • response illegible or incomprehensible
III. PERFORMANCE OF APPROPRIATE NEUTRALIZATION	• in a ***controlled manner***, either: 1) added vinegar to cleaner + indicator *or* 2) added cleaner to indicator + vinegar • noted ***single*** **drop-by-drop results**	• in a ***controlled manner***, either: 1) added vinegar to cleaner + indicator *or* 2) added cleaner to indicator + vinegar • noted ***something other than*** **single drop-by-drop results**	• did not use comparative ability to neutralize vinegar as basis of experiment (e.g., merely added cleaner *directly* to indicator)
IV. DRAWING CONCLUSIONS	• **correct** conclusion • conclusion is **consistent with** ***their*** **data**	• **incorrect** conclusion) • conclusion is **consistent with** ***their*** **data**	• correct or incorrect conclusion **BUT** • conclusion is **not consistent with** ***their*** **data**
V. EXPLANATION OF CONCLUSIONS	• response is **at least one** ***complete sentence*** that is **clear** and ***correctly*** **refers to their data**	• response is **clear** and correctly or incorrectly **refers to their data**	• **response is unclear** with or without reference to their data • response is incomprehensible or on another planet

Laboratory Network Program — *Alternative Assessment Toolkit*
Sample 3.1: *Acids-Bases Assessment Task*
LNP Contributor: Ann Muench, FWL

Permission for use in training has been granted by Far West Eisenhower Regional Consortium. Materials may be copied for training purposes only.

Self-Scoring Sheet Cleaning Up a Toxic Spill – Partner Task

Student Name(s) ______________________________

Color the Ós to show how well you think you did on each part of the task. Tell why.

[●ÓÓÓÓ = poor work ●●●ÓÓ = good work ●●●●● = excellent work]

(Space below is for teacher notes.)

How well we __did our experiment__ to find out which cleaners were bases.	ÓÓÓÓÓ because	*Determination of Bases*
How well we __told about what we did__ to find out which base was the strongest.	ÓÓÓÓÓ because	*Description of Experiment*
How well we __did our experiment__ to find out which base was the strongest.	ÓÓÓÓÓ because	*Performance of Appropriate Neutralization*
How well we __understood and used our results__ to decide which base was strongest.	ÓÓÓÓÓ because	*Drawing Conclusions*
How well we __explained our reasons__ why we knew which base was strongest.	ÓÓÓÓÓ because	*Explanation of Conclusions*

Permission for use in training has been granted by Far West Eisenhower Regional Consortium. Materials may be copied for training purposes only.

Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 3.1: *Acids-Bases Assessment Task*
LNP Contributor: Ann Muench, FWL

Permission for use in training has been granted by Far West Eisenhower Regional Consortium. Materials may be copied for training purposes only.

STUDENT NAMES ____ PARTNERS "A" ____

Cleaning Up a Toxic Spill

You are an Environmental Scientist. A truck has an accident on the highway and a barrel of a chemical called Big Messium spills on the road. The Highway Police ask for your help. You look up Big Messium in your computer and learn that the best way to clean it up is to use a strong base. You have four different cleaners that you are allowed to use to clean up highways. You need to do an experiment to decide which is the strongest base.

Supplies

Safety glasses – wear these when handling or near the solutions	Four labeled Cleaners in dropper bottles
Masking tape	Dropper bottles with pH Indicator and Vinegar
Ten Test tubes in test tube holder	

1) First you need to find out if any of the cleaners are a base. Describe in one or two sentences the simplest experiment that you can do to find out which of the cleaners are bases.

FIRST YOU PUT 7 DROPS OF INDICATOR AND ONE DROP OF VINEGAR. THEN YOU PUT IN A LIBERAL AMOUNT OF EACH CLEANER. IF IT CHANGES THE COLOR FROM PINK TO GREEN AND THEN BLUE IT IS A BASE.

2) Do the experiment and record your results in the Data Table.

Cleaner	Color Observations	Is It A Base?
A	IT TURNED FROM PINK TO YELLOW TO GREEN AND SO ON	YES
B	IT STAYED THE SAME PINK COLOR	NO
C	IT TURNED FROM PINK TO ORANGE TO GREEN AND SO ON	YES
D	IT TURNED FROM PINK TO ORANGE TO LIGHT GREEN AND SO ON	YES

3) If more than one cleaner is a base, you need to do an experiment to discover which is the strongest base. What materials are you going to use? Write step-by-step what you are going to do. Use a picture if that helps.

FIRST YOU PUT 7 DROPS OF INDICATOR AND ONE DROP OF VINEGAR AND PUT IN A LIBERAL AMOUNT OF A PARTICULAR BASE AND SEE HOW QUICK THE COLOR CHANGES TO PURPLE QUICKEST. THE QUICKEST SHOWS A STRONG BASE AND THE SLOWEST INDICATES A WEAK BASE

4) Do your experiment. Make a Data Table to show your results.

NUMBER OF DROPS	CLEANER A	CLEANER C	CLEANER D
DROPS 0	PINK	PINK	PINK
1	ORANGE	YELLOW	PINK
2	YELLOW	GREEN	ORANGE
3	LIGHT GREEN	TURQUOISE	LIGHT GREEN
4	BLUE GREEN	TURQUOISE	LIGHT GREEN
5	BLUE GREEN	LIGHT BLUE	LIGHT GREEN
6	TURQUOISE	BLUE	GREEN
7	TURQUOISE	BLUE	GREEN
8	TURQUOISE	BLUE-PURPLE	GREEN
9	LIGHT BLUE	BLUE-PURPLE	GREEN

5) Which of the cleaners is the strongest base?

CLEANER C IS THE STRONGEST BASE.

6) What evidence proves that it is the strongest base?

IT TURNED QUICKER TO PURPLE THAN THE OTHERS. NINE DROPS OF EACH IN THE TUBE WITH THE CLEANER C WAS BLUE-PURPLE. IN CLEANER A IT ONLY TURNED LIGHT BLUE AND IN CLEANER D IT TURNED GREEN.

BEST COPY AVAILABLE

Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 3.1: *Acids-Bases Assessment Task*
LNP Contributor Ann Muench, FWL

6

Permission for use in training has been granted by Far West Eisenhower Regional Consortium. Materials may be copied for training purposes only

STUDENT NAMES ___ PARTNERS "B" ___

Cleaning Up a Toxic Spill

You are an Environmental Scientist. A truck has an accident on the highway and a barrel of a chemical called Big Messium spills on the road. The Highway Police ask for your help. You look up Big Messium in your computer and learn that the best way to clean it up is to use a strong base. You have four different cleaners that you are allowed to use to clean up highways. You need to do an experiment to decide which is the strongest base.

Supplies

- Safety glasses – wear these when handling or near the solutions
- Masking tape
- Ten Test tubes in test tube holder
- Four labeled Cleaners in dropper bottles
- Dropper bottles with pH Indicator and Vinegar

1) First you need to find out if any of the cleaners are a base. Describe in one or two sentences the simplest experiment that you can do to find out which of the cleaners are bases.

ADD INDICATER THEN CLEANER. THEN ADD VINEGAR

2) Do the experiment and record your results in the Data Table.

Cleaner	Color Observations	Is It A Base?
A	PURPLE + VINEGAR = YELLOW	YES
B	PINK + VINEGAR = PINK	NO
C	PURPLE + VINEGAR = GREEN	YES
D	PURPLE + VINEGAR = ORANGE	YES

3) If more than one cleaner is a base, you need to do an experiment to discover which is the strongest base. What materials are you going to use? Write step-by-step what you are going to do. Use a picture if that helps.

USE SAFETY GLASSES, VINEGAR, INDICATER, CLEANER A, C, AND D.

ADD INDICATER, THEN THE CLEANERS. IF THE COLOR CHANGES THE MOST WHEN YOU ADD THE VINEGAR IT'S THE STRONGEST BASE

4) Do your experiment. Make a Data Table to show your results.

	CLEANER A	CLEANER C	CLEANER D
5 DROPS	YELLOW	GREEN	ORANGE
10 DROPS	PINK	PEACH	PINK
0 DROPS	PURPLE	PURPLE	PURPLE

5) Which of the cleaners is the strongest base?

CLEANER C

6) What evidence proves that it is the strongest base?

THE COLOR CHANGE IN C IS MORE POWERFUL (GREEN TO PEACH IS FURTHER THAN YELLOW TO PINK AND ORANGE TO PINK).

BEST COPY AVAILABLE

Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 3.1: *Acids-Bases Assessment Task*
LNP Contributor: Ann Muench, FWL

Permission for use in training has been granted by Far West Eisenhower Regional Consortium. Materials may be copied for training purposes only.

STUDENT NAMES ___ PARTNERS "C" ______

Cleaning Up a Toxic Spill

You are an Environmental Scientist. A truck has an accident on the highway and a barrel of a chemical called Big Messium spills on the road. The Highway Police ask for your help. You look up Big Messium in your computer and learn that the best way to clean it up is to use a strong base. You have four different cleaners that you are allowed to use to clean up highways. You need to do an experiment to decide which is the strongest base.

Supplies

Safety glasses – wear these when handling or near the solutions	Four labeled Cleaners in dropper bottles
Masking tape	Dropper bottles with pH Indicator and Vinegar
Ten Test tubes in test tube holder	

1) First you need to find out if any of the cleaners are a base. Describe in one or two sentences the simplest experiment that you can do to find out which of the cleaners are bases.

The one that turns purple is a base

2) Do the experiment and record your results in the Data Table.

Cleaner	Color Observations	Is It A Base?
a	the color purple with one drop	yes
b	the color peach	no
c	the color purple	yes
d	the color purple	yes

3) If more than one cleaner is a base, you need to do an experiment to discover which is the strongest base. What materials are you going to use? Write step-by-step what you are going to do. Use a picture if that helps.

use indicator and a, b, and c
amonia vingar
Test each one out which ever one turns color first is the strongest

4) Do your experiment. Make a Data Table to show your results.

	number of drops	what colors
A	1	purple
B	0	peach
C	5	purple
D	2	Dark purple

5) Which of the cleaners is the strongest base? D

6) What evidence proves that it is the strongest base?

it changed color with the less drops

BEST COPY AVAILABLE

Sample 4.1

How Many Buttons?

Mathematical Sciences Education Board

This exercise is one of 13 in *Measuring Up*, designed to illustrate performance assessment tasks that could be used with fourth graders to support reform efforts in mathematics instruction. These prototypes are meant to be examples of assessment possibilities, not examples of ready-to-use assessments.

*A Task in: "Measuring Up—Prototypes for Mathematics Assessment," 1993**
Mathematical Sciences Education Board of the National Academy of Sciences
National Academy Press
2101 Constitution Ave. NW
Washington, DC 20418
(800) 624-6242

*Copies of the book are available from the National Academy Press at $10.95 per copy. Quantity discounts are available.

Permission for use in training has been granted by the Mathematical Sciences Education Board. Materials may be copied for training purposes only.

Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 4.1: *How Many Buttons?*
LNP Contributor: Judy Arter, NWREL

Permission for use in training has been granted by the Mathematical Sciences Education Board. Materials may be copied for training purposes only.

How Many Buttons?

Use approximation and estimation

Use calculators appropriately

Link assessment and instruction

Suggested time allotment
A total of two or three class periods (one class period of introductory work plus one or two class periods of student work)

Student social organization
Students working in groups of two or three

Task

Assumed background: This task assumes that the children have had experience in gathering data and making estimates in relatively complex situations.

Presenting the task: This assessment is embedded in a context-setting activity introduced by the teacher to the whole class. The context-setting activity takes one class period, and the paired work—the assessment task itself—requires a second class period, possibly spilling over into part of a third.

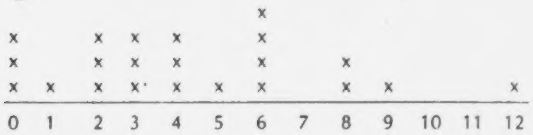
The following is written as directions to the teacher:

Introduce the activity by telling the students they will be working on estimating the number of buttons *in their class* today.

Ask students for their estimates of the number of buttons in the class. Give students a few minutes to look around—at their own buttons, at the number of buttons that students near them have—and to discuss their estimates with students near them. Collect some of their estimates and list them on the board.

As students are estimating, some questions and issues will probably arise. Encourage students to talk about any questions that occur to them (for example, "What about buttons on sweaters or coats hanging in the closet?" "Were there buttons that they didn't notice or forgot about because they were not in obvious places?" "Should the teacher's buttons be included?" The answers to these questions, which can be decided jointly by you and your students, will depend on your particular situation. For instance, if students' outerwear is stored in an inconvenient location you might decide not to include it.)

Now each student should carefully count the number of buttons that he or she has. They might work in pairs or threes to help each other count carefully. Record their data on a line plot so that everyone in the class can see it. For example:



Encourage students to describe these data by asking: "What would you say about the number of buttons in our class?" Students may comment on the range of the data, where data are clumped, for

BEST COPY AVAILABLE

Permission for use in training has been granted by the Mathematical Sciences Education Board. Materials may be copied for training purposes only.

which values there are no people with exactly that number of buttons, and what (if anything) seems typical of their class.

Student assessment activity: The assessment should be done on the day after the activity just described. You will need the class data from the previous day, and pencil and paper for each student. Of course, they should have access to calculators.

Tell students they will be working in small groups to estimate the number of buttons there are in the whole school. They can refer to the class data from the previous session. Suggest that they may want to consider questions such as: "Is it likely that other classes in the school will have a similar number of buttons?" "Do younger students have fewer or more buttons?" "What about older students?" "How should teachers or other adults in the school be included in the estimate?" These questions should be written on the board. They can contribute greatly to the depth and sophistication of the children's responses, so students should have equal access to them.

Each pair of students then works together to solve this problem, as described above. They should be encouraged to record their strategies, including the numbers they chose and why they chose them.

Rationale for the mathematics education community

This item pushes the curriculum to include work with large numbers and with a complex situation in which estimation is a *legitimate* mathematical process—not simply a prelude to finding the "real" answer. Further, it connects a range of mathematical ideas—multiplicative relationships (e.g., an average of about 6 buttons per student, with 22 students per classroom), estimation, averages ("I decided there are about 22 students in each classroom."), and beginning ideas about sampling. Use of the calculator is incorporated in a natural way.

The problem is accessible: Almost any child can arrive at some solution, but the possible solutions span a broad spectrum of depth and complexity. The problem provides all students with the opportunity to think about and discuss which factors to take into account — whether or not to include adults' buttons or clothes in the closet, and so forth. One strength of a task like this is that its context is immediate and tangible.

Many estimation tasks arise in some setting that provides a "real-life" motivation and a criterion for judging how precise an estimate should be. One must remember, however, that fourth graders are not as concerned as adults are with the "real" reasons for doing something. In this case, classroom trials have suggested that just the allure of the large numbers, of finding out something that no one has ever known before, and of getting the best possible estimate, are plenty of motivation.

The mode of presentation embeds the assessment task as an integral part of a piece of curriculum. Indeed, teachers might want to move from the assessment task back into curriculum by continuing the investigation with the whole class—comparing the students' individual estimates, gathering additional information, and refining these estimates. (See the Variants and extensions section below.)

Finally, it offers a problem on which students work in pairs. Working together to pool information and expertise to come to a solution are essential problem-solving skills for today's students.

Task design considerations: The protorubric for this task is phrased in terms of the factors that the child's response takes into account. Nonetheless, careful consideration must be given to the specific details of the task if it is to be used to make comparisons among, or to draw conclusions about, different kinds of schools in different geographical areas of the country. Depending on the season, a child in a colder climate is likely to wear more clothes, with more buttons, than one from a warmer climate. A school that contains 25 classrooms is a very different setting from one that has five classrooms. Because of such potential variations in complexity, the task would be inappropriate to use universally for comparative purposes.

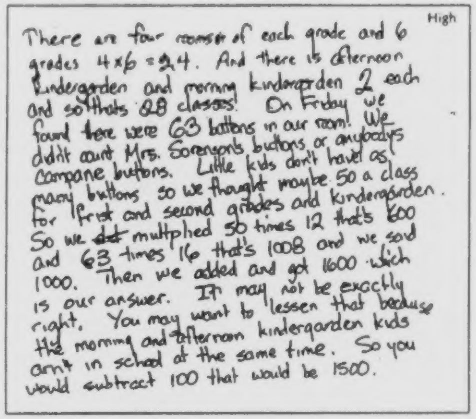
Variants and extensions: Objects other than buttons can be used in this sort of activity — for example, books that the student happens to have in his or her desk. In fact, this might alleviate (but not eliminate) some of the geographical differences noted above.

Using the task as the basis for extended classroom discussion might be particularly valuable. One can continue the investigation by considering how many buttons there might be in all the schools in the town or city, or even all the schools in the state. This would involve the use of almanacs or atlases to find out the appropriate numbers to use, depending on the question. Similarly, students might explore how the number of buttons that people wear varies from one section of the country to another. Thus, the students would want to communicate with at least one other school of the same size in some other region and compare data.

Protorubric

Characteristics of the high response:

The response appropriately takes into consideration the following aspects of the problem: (a) the data collected from the day before; (b) some estimate of the number of children per classroom; (c) the number of classes in the school; (d) variation of buttons among children of various ages; and (e) adults' buttons (which the pair may choose to disregard entirely).


High

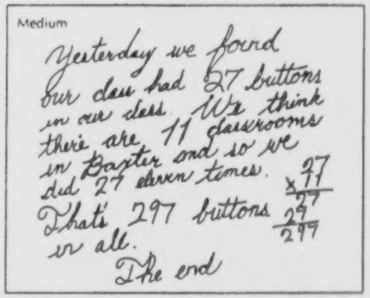
There are four rooms of each grade and 6 grades 4×6=24. And there is afternoon kindergarden and morning kindergarden 2 each and so thats 28 classes. On Friday we found there were 63 buttons in our room! We didn't count Mrs. Sorenson's buttons or anybodys compane buttons. Little kids don't have as many buttons so we thought maybe 50 a class for frist and second grades and kindergarden. So we ~~did~~ multplied 50 times 12 that's 600 and 63 times 16 that's 1008 and we said 1000. Then we added and got 1600 which is our answer. It may not be exactly right. You may want to lessen that because the morning and afternoon kindergarden kids arn't in school at the same time. So you would subtract 100 that would be 1500.

Reasonable justifications are given for how each numerical value was chosen or calculated.

The numerical values are put together with appropriate arithmetic processes. (For example, if the students decided there were an average of 90 buttons in each of the 12 upper grade classrooms and an average of 60 buttons in each of the 8 lower grade classrooms, they would calculate with these values so that their result would be (90 x 12) + (60 x 8), not in some invalid way, such as adding 90 and 60, then multiplying by the total number of classrooms.)

The final report on the solution clearly details the steps the students went through to solve the problem so that a reader can easily follow their solution strategy.

Characteristics of the medium response:


Medium

Yesterday we found our class had 27 buttons in our class. We think there are 11 classrooms in Baxter and so we did 27 eleven times.
That's 297 buttons in all.
The end

$$\begin{array}{r} 27 \\ \times 11 \\ \hline 27 \\ 27 \\ \hline 297 \end{array}$$

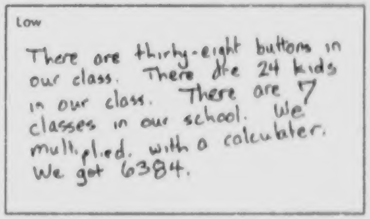
The response takes into consideration the results of the previous day's work. Further, it uses a reasonable estimate for the number of students per class (if needed for the solution) and the number of classrooms in the school, with appropriate justifications for the values chosen.

Consideration of factors like variation of button numbers by age of child or adults' buttons is lacking.

The arithmetic processes are appropriate.

The final report on the solution clearly explains some (but not all) aspects of the student's strategy.

Characteristics of the low response:


Low

There are thirty-eight buttons in our class. There are 24 kids in our class. There are 7 classes in our school. We multiplied with a calculater. We got 6384.

Some reference is made to the previous day's work, and some other values may be chosen reasonably.

Either some values (e.g., the number of students per class or the number of classes) are chosen inappropriately, or the way in which the arithmetic processes are used is incorrect.

The explanation and justification of the student's reasoning is incomplete or unclear.

Reference

"Investigations in Number, Data, and Space" project of TERC, Cambridge, MA.

Sample 4.2

MATHEMATICS PROBLEM SOLVING

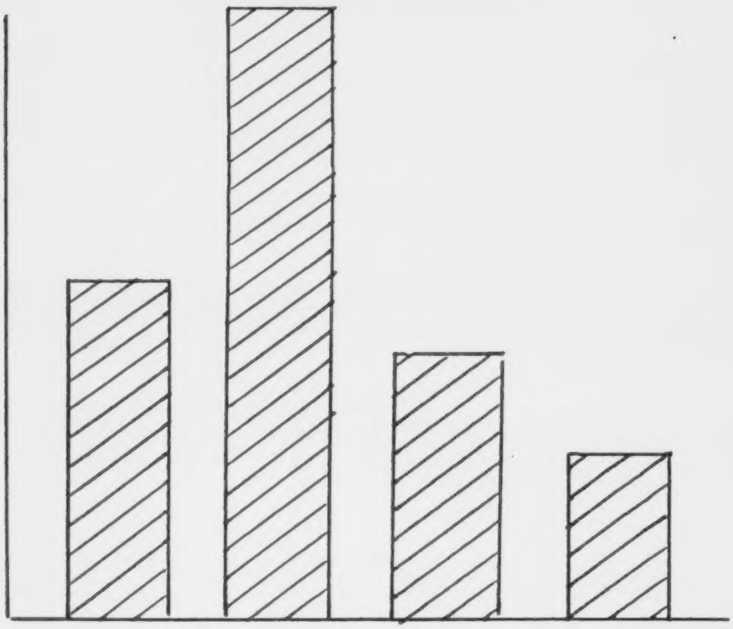
OREGON DEPARTMENT OF EDUCATION

The Oregon State Assessment measures math problem solving in grades 3, 5, 8, and 11. Students are given short open-ended questions to solve individually. Students write down their answers and solution strategies. All papers are scored twice using the same generic scoring guides. Attached is a sample grade 5 problem and the rubric.

"Mathematics Problem Solving Scoring Guide," 1994
Oregon Department of Education
Office of Assessment
Public Services Building
255 Capitol St. NE
Salem, Oregon 97310-0203
(503) 378-8004

Permission for use in training has been granted by the Oregon Department of Education, 255 Capitol NE, Salem, OR 97310. Materials may be copied for training purposes only.

NAME THE GRAPH



1. **What do you think this might be the graph of? Put names and numbers on the graph to show what you mean.**
2. **Write down everything you know from your graph.**

Permission for use in training has been granted by the Oregon Department of Education, 255 Capitol NE, Salem, OR 97310. Materials may be copied for training purposes only.

MATHEMATICS PROBLEM SOLVING SCORING GUIDE

Dimensions of Mathematics

CONCEPTUAL UNDERSTANDING: Scoring Guide

Conceptual understanding includes the ability to interpret the problem and select appropriate information to apply a strategy for solution. Evidence is communicated through making connections between the problem situation, relevant information, appropriate mathematical concepts and logical/reasonable responses.

5 **FULL CONCEPTUAL UNDERSTANDING: The student uses all relevant information to solve the problem.**
- The student's answer is consistent with the question/problem.
- The student is able to translate the problem into appropriate mathematical concepts.

3 **PARTIAL CONCEPT UNDERSTANDING: The student extracts the "essence" of the mathematics in the problem, but is unable to use this information to solve the problem.**
- The student is only partially able to make connections between/among the concepts.
- The student's solution is not fully related to the question.
- The student understands one portion of the task, but not the complete task.

1 **LACK OF CONCEPTUAL UNDERSTANDING: The student's solution is inconsistent or unrelated to the question.**
- The student translates the problem into inappropriate mathematical concepts.
- The student uses incorrect procedures without understanding the concepts related to the task.

PROCEDURAL KNOWLEDGE: Scoring Guide

Procedural Knowledge deals with the student's ability to demonstrate appropriate use of concepts. Evidence includes the verifying and justifying of a procedure using concrete models, or the modifying of procedures to deal with factors inherent in the problem.

5 **FULL USE OF APPROPRIATE PROCEDURES: The student uses principles efficiently while justifying the solution.**
- The student uses appropriate mathematical terms and strategies.
- The student solves and verifies the problem.
- The student uses mathematical principles and language precisely.

3 **PARTIAL USE OF APPROPRIATE PROCEDURES: The student is not precise in using mathematical terms, principles, or procedures.**
- The student is unable to carry out a procedure completely.
- The process the student uses to verify the solution is incorrect.

1 **LACKS USE OF MATHEMATICAL PROCEDURES: The student uses unsuitable methods or simple manipulation of data in his/her attempted solution.**
- The student fails to eliminate unsuitable methods or solutions.
- The student misuses principles or translates the problem into inappropriate procedures.
- The student fails to verify the solution.

Permission for use in training has been granted by the Oregon Department of Education, 255 Capitol NE, Salem, OR 97310. Materials may be copied for training purposes only.

PROBLEM SOLVING STRATEGIES: Scoring Guide

Problem Solving requires the use of many skills, often in certain combinations, before the problem is solved. Students demonstrate problem solving strategies with clearly focused, good reasoning that lead to a successful resolution of a problem.

5 **EVIDENCE OF THOROUGH/INSIGHTFUL USE OF SKILLS/STRATEGIES: The skills and strategies show some evidence of insightful thinking to explore the problem.**

- The student's work is clear and focused.
- The skills/strategies are appropriate and demonstrate some insightful thinking.
- The student gives possible extensions or generalizations to the solution or the problem.

3 **EVIDENCE OF ROUTINE OR PARTIAL USE OF SKILLS/STRATEGIES: The skills and strategies have some focus but clarity is limited.**

- The student applies a strategy which is only partially useful.
- The student's strategy is not fully executed.
- The student starts the problem appropriately, but changes to an incorrect focus.
- The student recognizes the pattern or relationship, but expands it incorrectly.

1 **LIMITED EVIDENCE OF SKILLS/STRATEGIES: The skills and strategies lack a central focus and the details are sketchy or not present..**

- The procedures are not recorded (i.e. only the solution is present).
- Strategies are random.
- The student does not fully explore the problem looking for concepts, patterns or relationships.
- The student fails to see alternative solutions that the problem requires.

COMMUNCATION: Scoring Guide

In assessing the student's ability to communicate mathematically, particular attention should be paid to both the meanings he/she attaches to the concepts and procedures and also his/her fluency in explaining, understanding, and evaluating the ideas expressed.

5 **CLEAR, COMPLETE COMMUNICATION: The student gives a complete response with clear, coherent, unambiguous, and elegant explanations.**

- The student communicates his/her thinking effectively to the audience.
- The details fit and to make sense.
- One step flows to the next and shows organization.
- The student presents strong supporting arguments.

3 **PARTIAL OR INCOMPLETE COMMUNICATION: The student's explanation is unclear, inconsistent or not complete.**

- The student uses terminology incorrectly or inconsistently.
- The student's visual aids (graphs, tables, diagrams, etc.) are inappropriate, or not directly related.
- The student's explanation centers on his/her solution, not on his/her thinking.

1 **LIMITED OR LACK OF COMMUNICATION: The student's explanation is not understandable or not present.**

- The student either does not use or misuses appropriate mathematical terminology.
- The student does not use essential visual aids to enhance or clarify the explanation.
- The student's explanation lacks focus.

Permission for use in training has been granted by the Oregon Department of Education, 255 Capitol NE, Salem, OR 97310. Materials may be copied for training purposes only.

Sample 4.3

DISCOVERING THE PROBLEM OF SOLID WASTE

LAKE COUNTY EDUCATIONAL SERVICE CENTER, ILLINOIS

Teachers developed 17 performance tasks for students in grades 3-6 to assess their knowledge of solid waste and recycling, and ability to apply their knowledge to real-life tasks. Tasks cover such things as concepts (e.g., types of solid waste, changes in materials in a landfill), and process skills (e.g., classifying, observing).

Attached is one of the tasks and the associated scoring instructions.

Dr. Elaine P. Lee, Project Director
Dr. Jerry Van Pelt, Manager
"Discovering the Problem of Solid Waste," 1991
Lake County Educational Service Center
19525 W. Washington St.
Grayslake, IL 60030
(708) 223-3400

Permission for use in training has been granted by the Lake County Educational Service Center. Materials may be copied for training purposes only.

TEMPLATE #1

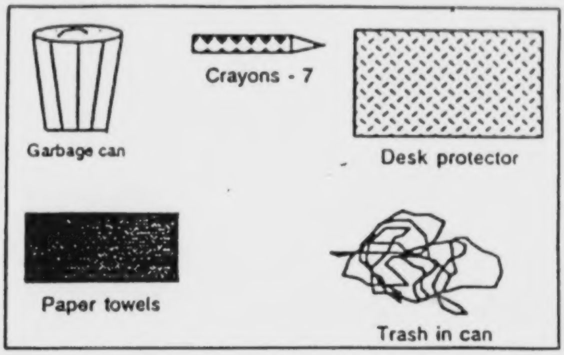
Name:
Date: 1/18/92
Grade Level: 4, 5
School:

KEY - Grade Level
Concept (s): Types of Solid Waste
Process Skill (s): Classification, Ordering, Graphing, Hypothesizing

STATION #: One
Do you have the items in the diagram?

1. Examine the items in the garbage can.
2. Sort the items into the following categories:

Paper	Plastic
Metal	Yard Waste
Food Waste	Glass
Other	



3. Count and record the number of items in each category.

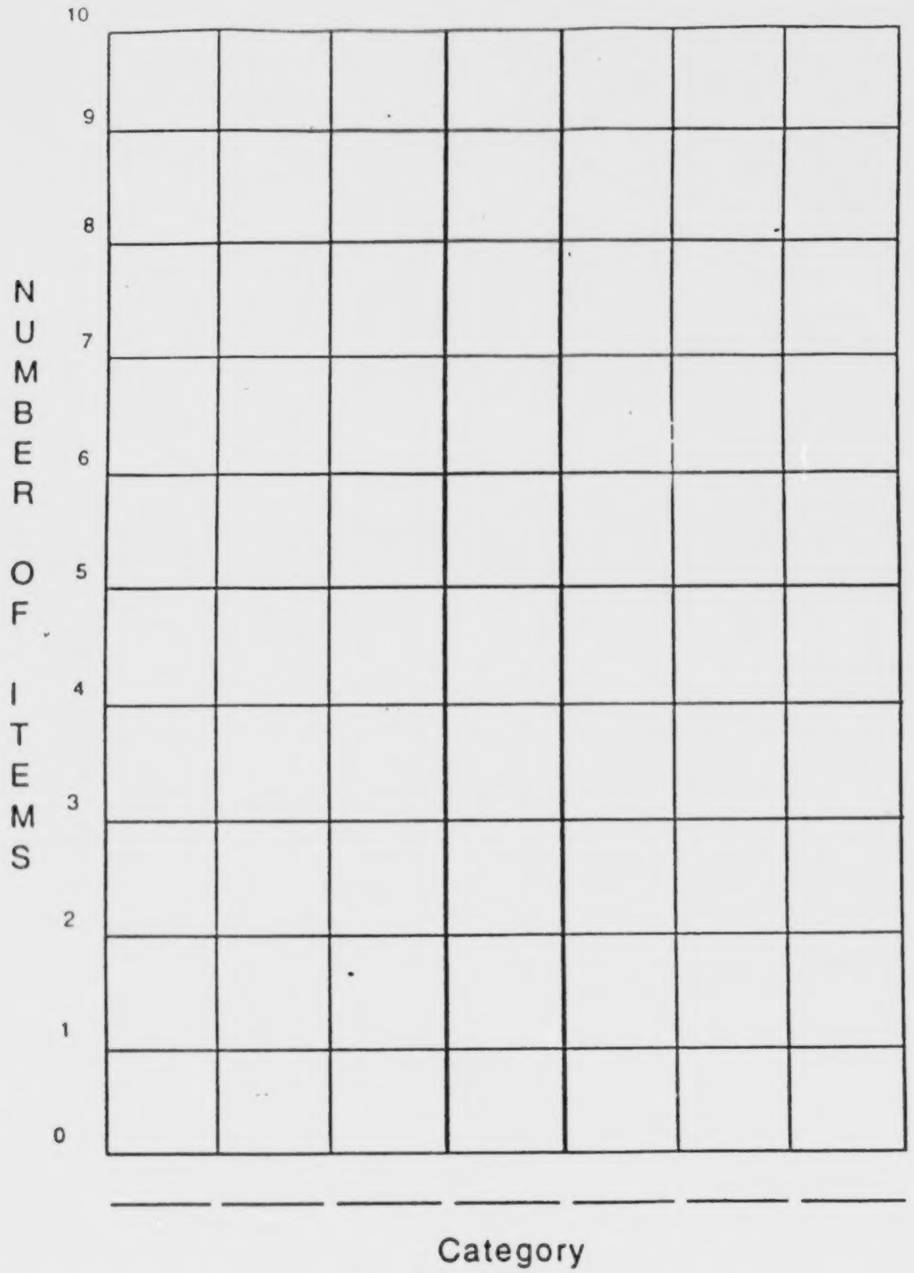
Paper _______
Plastic _______
Other _______
Metal _______
Yard waste _______
Food waste _______
Glass _______

4. List the categories from above in order from least number of items to greatest number of items.

Least ____________________
____
____
____
____
Greatest ____________________

5. Use your answers from questions #3 and #4 to help complete the graph on page two. Place the names of each category on the lines below the graph in order from least to greatest.

Permission for use in training has been granted by the Lake County Educational Service Center. Materials may be copied for training purposes only.



6. Color the data for each category a different color.

7. What could you do in order to reduce the amount of solid waste in this garbage can?

____

____

TEMPLATE #2

Discipline Area: Science

Goal: ISBE Goal: 2: Social and Environmental Implications. Objective: F: application of selected ecological concepts to human and environmental situations. ISBE Goal 4: Processes, techniques. Objectives: Observation (a) classification, (b) data collection, organization, interpretation (g), graphing (e).

Valued Outcome: The students will be able to correctly classify solid waste into the appropriate categories.
The students will be able to graph the results of their data collection.
The students will be able to numerically order the data results.

Assessment Category: Physical Science

Mode of Presentation: Performance Assessment tool to be completed by individual student.

Technology Requirements (Materials):

- 1 plastic trash can
- 1 desk protector
- 7 crayons - different colors
- paper towels
- trash items:
  - paper - 10 items
  - plastic - 7 items
  - metal - 6 items
  - yard waste - 4 items
  - food waste - 3 items
  - glass - 1 item
  - others - 5 items

Description of Assessment Task:

1. Obtain assessment tool from teacher.
2. Go to open station.
3. Read the directions at the station.
4. Complete the task.
5. Return to your seat and complete the rest of the task.

Performance Criteria:
Rubric

3. Sorting
   7 points for all correct
   Subtract 1 point for each incorrect response

4. Sequencing
   1 point for ordering items correctly
   0 points for incorrect order

5. Graph
   2 points for labeling and ordering categories on horizontal axis
   - subtract 1 point for incorrect order
   7 points for correct graphic representation of each category
   - subtract 1 point for each incorrect representation

6. Solution
   4 points for a strong, workable solution
   3 points for solution that is unworkable
   1 point for any attempt

Permission for use in training has been granted by the Lake County Educational Service Center. Materials may be copied for training purposes only.

Sample 4.4

Sow Bugs

University of California at Santa Barbara

Students receive five sow bugs, a round dish to contain them, a bright light and strips of dark cardboard to create regions of light and dark, filter paper and a spray bottle for creating damp regions, and a stopwatch. Students are to answer the following questions:

1. Do sow bugs prefer light or dark environments?
2. Do sow bugs prefer damp or dry environments?
3. Do dampness and amount of light in combination make a difference on sow bug preferences?

Scoring was procedure based. For each experiment, observers focused on the method used to solve the problem (e.g., two conditions in a dish or one condition at a time), the adequacy with which conditions were manipulated (e.g., equal area for each condition), the measurement strategies used to determine the result (e.g., count the number of bugs in each condition), and the correctness of the solution generated.

Sources of information:

Maria Araceli Ruiz-Primo, Gail P. Baxter, and Richard J. Shavelson. "On the Stability of Performance Assessments," *Journal of Educational Measurement, Vol. 30,* Spring 1993, pp. 41-51.

Richard J. Shavelson, Gail P. Baxter, and Jerry Pine. "Performance Assessments: Political Rhetoric and Measurement Reality," *Educational Researcher,* May 1992, pp. 22-27.

Jerome Pine, Gail Baxter, and Richard Shavelson. *Assessments for Hands-On Elementary Science Curricula,* Graduate School of Education, University of California at Santa Barbara, Santa Barbara, CA 93106, 1991.

Permission for use in training has been granted by Richard J. Shavelson. Materials may be copied for training purposes only.

Sample 4.5

Weathercaster's Helper

Utah Department of Education

The Utah Department of Education has developed 30 constructed response items in science for grades 1-6 to complement the multiple-choice tests already in place. These are model assessments, not required assessments. Districts must assess student status with respect to the Utah Core Curriculum goals, but the use of these tests for that purpose is voluntary.

Each of the five scenarios at each grade level contains a series of questions related to the theme of the scenario. The questions are designed to measure four general process skills: identify/describe, explain/infer, organize, and create. Points for each scenario are totaled. Then points are added up between questions for each of the skill areas being assessed. Four levels of proficiency on each skill (advanced, proficient, basic, and below basic) are identified depending on percent of total possible points.

The attached samples are a portion of one of the five activities and related scoring guide for grade 1.

Dr. Barbara Lawrence
Utah Core Curriculum Performance Assessment Program: Science, 1993
Profiles Corporation
507 Highland Avenue
Iowa City, IA 52240

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

(Discuss what each weather symbol refers to and how the symbols stand for the different kinds of weather that a person can observe. Also discuss how symbols are an easy way for a weathercaster to record the weather.)

1. **The weather on Saturday was different than the other days. It changed during the day. How would you show the weather on Saturday with weather symbols if it was sunny in the morning, then it rained in the afternoon?** (Repeat as necessary. Point out the enlarged box for Saturday and instruct students to draw the weather symbols in this box. Allow students time to complete the task.)

ACTIVITY II

Say: **Weathercasters record the weather every day. Your next job will be to record today's weather.** (Have your students observe the weather today. If you don't have a window, take them outside for a few minutes.)

Turn to page 4, the page with the boat at the top.

2. **Draw a picture of what the weather looks like today in the "window" on the top half of the page. Make your picture as complete as you can.** (Allow students reasonable time to complete their picture.) **If you want to keep working on your drawing, you may come back to it later.**
3. **Now look at the weather symbols at the bottom of the page. Choose the weather symbol that best describes the weather today and draw a circle around it. The picture you drew in the window and the weather symbol you will circle are related: Both are a way of describing and recording today's weather.** (Pause to allow students time to complete the task.)

ACTIVITY III

Say: **Turn to the last page of your booklet, the page with the pictures of trees.**

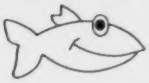
You will find four boxes with the names of the seasons in them. (Write the names of the seasons on the board and say the words aloud as you write them.) **Cut them out and then set them aside on your desk.** (Pause to allow them to cut out the boxes.)

Weathercasters know that the weather changes with the seasons. As the weather changes, there are also changes in the plants. Notice that in each box there is a picture that shows a tree you might see during that season. You can help the weathercaster by arranging the names of the seasons in the order that they come each year.

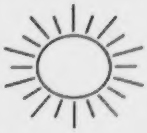
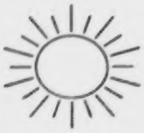
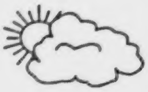
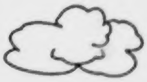
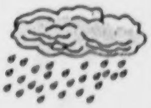
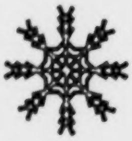
Now turn to page 5, the page with the balloon at the top.

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

ACTIVITY I



Weekly Weather

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday
					

Saturday

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

SCORING GUIDE

Activity I

1. **3 pt.** Student draws both sunny and rainy weather symbols in any arrangement that indicates it was sunny first and rainy second.

2 pt. Student draws both sunny and rainy weather symbols but their arrangement does not indicate the order in which they occurred.

1 pt. Student only begins to address the task; for example, draws either the sunny or rainy weather symbol but not both.

0 pt. Student does not answer or answers inappropriately.

Activity II

2. **2 pt.** Student draws a detailed and accurate picture of the day's weather.

1 pt. Student only begins to address the task; for example, draws a cloud.

0 pt. Student does not answer or answers inappropriately.

3. **1 pt.** Student circles a weather symbol that accurately symbolizes the day's weather.

0 pt. Student does not answer or answers inappropriately.

Activity III

4. **2 pt.** Student arranges the seasons in the correct order, may begin with any season in the top box.

1 pt. Student only begins to address the task; for example, arranges two of the four seasons in the correct order.

0 pt. Student does not answer or answers inappropriately.

Activity IV

5. **3 pt.** Student writes the name of a season and dresses the figure appropriately for that season in the student's local area. Dress is complete from head to toe (though barefoot is appropriate for warm weather).

2 pt. Student addresses the task but something is missing; for example, draws clothes on the figure that is typical for a certain season but does not write the name of the season.

1 pt. Student only begins to address the task; for example, writes the name of a season but does not dress the figure, or draws only one or two articles of clothing on the figure.

0 pt. Student does not answer or answers inappropriately.

Total points possible: 11

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

Sample 4.6

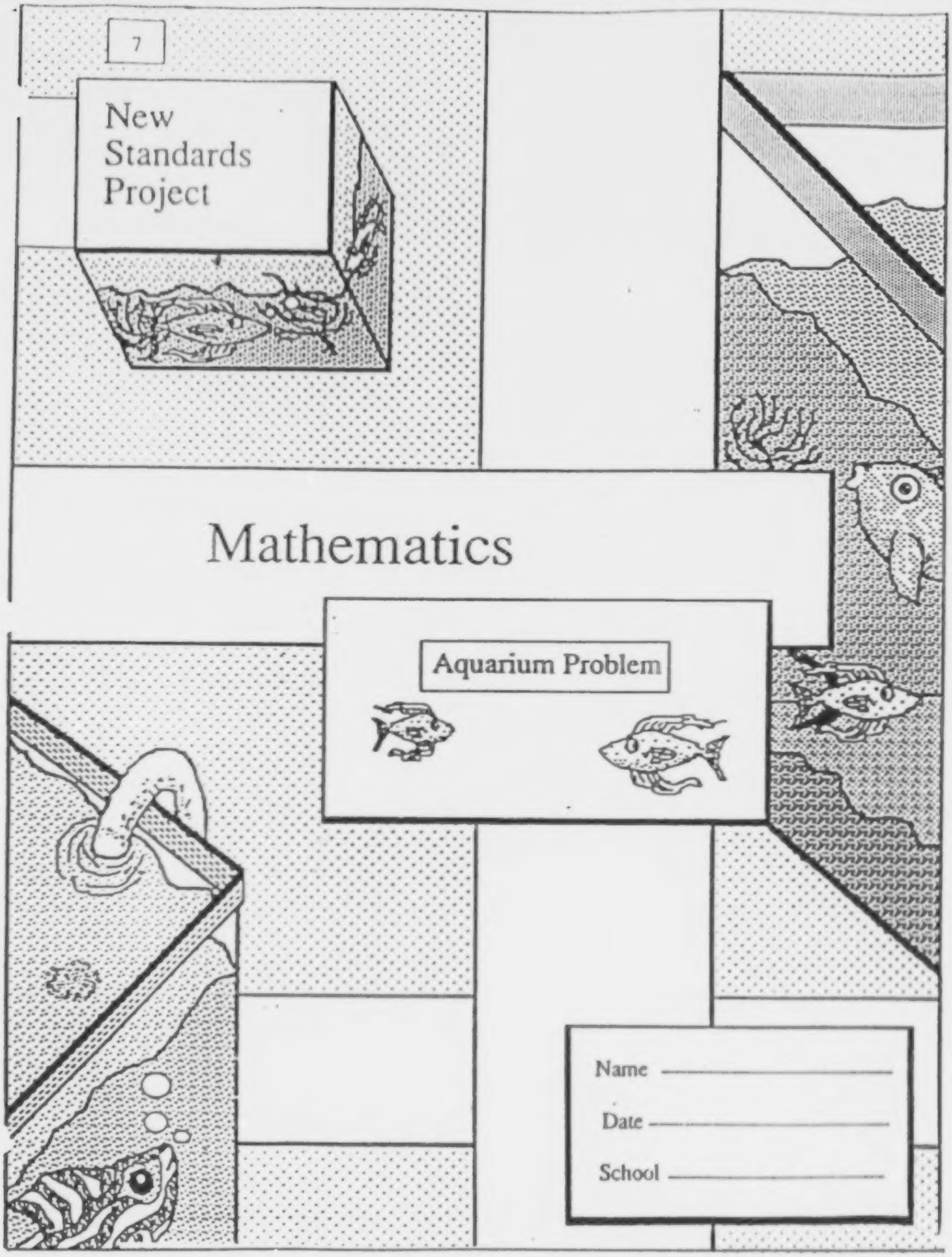
Aquarium Problem

New Standards Project

This problem is one of several developed by the New Standards Project at the University of Pittsburgh. Attached are the complete instructions for the problem, the scoring guide, and five student responses representing different levels of performance.

"Aquarium Problem," undated
New Standards Project
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 18260

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.



7

New Standards Project

Mathematics

Aquarium Problem

Name ____________________

Date ____________________

School ____________________

426

THE AQUARIUM

Imagine that your school principal asks you to do a special job and gives you these written directions:

> Your class will be getting a 30 gallon aquarium. The class will have $25.00 to spend on fish. You will plan which fish to buy. Use the Choosing Fish for Your Aquarium brochure to help you choose the fish. The brochure tells you things you must know about the size of the fish, how much they cost and their special needs.
>
> Choose as many different kinds of fish as you can. Then write a letter to me explaining which fish you choose. In your letter,
>
> 1. tell me how many of each kind of fish to buy
> 2. give the reasons you chose those fish
> 3. show that you are not overspending and that the fish will not be too crowded in the aquarium.

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 4.6: *Aquarium Problem*
LNP Contributor: Judy Arter, NWREL

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

Choosing Fish for Your Aquarium

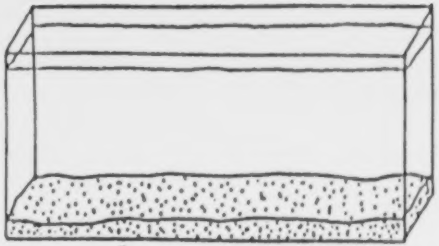
Planning Ahead

Use the information in this brochure to help you choose fish that will be happy and healthy in your aquarium. To choose your fish, you must know about the size of the fish, their cost, and their special needs.

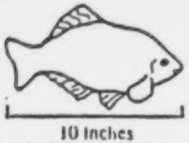
Size of Fish

To be healthy, fish need enough room to swim and move around. A good rule is to have one inch of fish for each gallon of water in your aquarium. This means that in a ten gallon aquarium, the lengths of all your fish added up can be ten inches at the most.

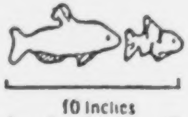
EXAMPLE:
With a ten gallon aquarium,



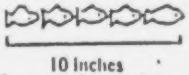
here are a few of your choices:



one ten-inch long fish, or



a seven-inch long fish and a three-inch long fish or



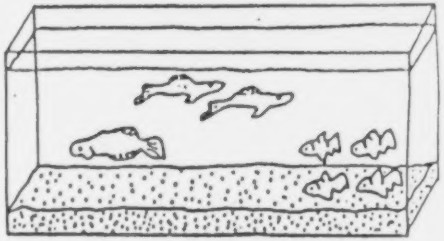
five fish if each is only two inches long.

Cost of the Fish

Some fish cost as little as one dollar, others cost much more. The prices of each kind of fish are listed in the chart.

Special Needs

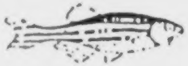
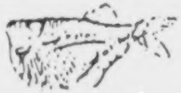
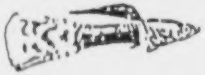
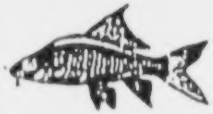
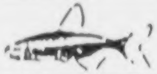
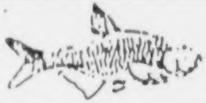
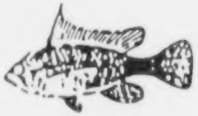
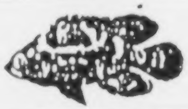
Use the chart to learn about the special needs of each kind of fish. Some fish need to live together in schools --a group of four or more of the same kind of fish -- while other live in pairs or alone. A few kinds of fish have other special needs, which are listed in the chart.



Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 4.6 *Aquarium Problem*
LNP Contributor: Judy Arter, NWREL

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only

Chart for Freshwater Fish

Picture	Name	Cost	Length In Inches	Color	Special Needs, Facts
	Zebra Danio	$1	$1\frac{1}{2}$ inches	blue with gold lines	Lives in schools; gets along with other kinds of fish.
	Marbled Hatchetfish	$1	2 inches	yellow	Lives in schools; can leap 3-5 yards.
	Guppy	2 for $3	2 Inches	red, blue and green	Lives in schools.
	Red-tailed Black Shark	$5	$4\frac{1}{2}$ inches	black with red tail	Fights with other sharks, but gets along with other kinds of fish.
	Cardinal Tetra	$5	$1\frac{1}{2}$ Inches	red and green	Lives in schools.
	Blind Cave Fish	$2	3 Inches	silvery rose	Lives in schools; uses its sense of smell and vibration to find food.
	Ramirez' Dwarf Cichlid	$5	2 Inches	rainbow	Lives in pairs; rarely lives longer than 2 $\frac{1}{2}$ years; gets along with other fish.
	Velvet Cichlid	$5	12 $\frac{1}{2}$ Inches	olive with stripes	Can be trained to take food from the hand and can be petted. Must be kept only with other cichlids

Student Reflections, Ideas

You can help us make these learning activities even better. Think about each of the following questions, and write to us what you honestly think. Be as clear as you can (you might want to give us examples of what you mean).

What did you enjoy about the task?

What did you not like about the task?

How is this task like other activities you do in your class? How is it different?

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

Aquarium Scoring Guide

Purpose

In the Aquarium task, students use their knowledge of mathematics to solve a real world problem: Students use logical and numerical reasoning about money, measurement and realistic conditions to decide how best to stock an aquarium within the constraints of the situation.

This task calls for logical and numerical reasoning and justification of that reasoning. These are contained in Standard 1 (Mathematics as Problem Solving), Standard 3 (Mathematics as Reasoning), and Standard 4 (Mathematical Connections) of the *NCTM Curriculum Standards* for Grades K-4. Students apply understanding of measurement as described in Standard 10 (Measurement).

This task is designed to assess students' mathematical thinking and their use of information contained in a brochure which features a chart, not their ability to read the brochure and chart independently.

On-Demand Performance

"On-demand" performance refers to performance "demanded" at a specific time and place under controlled circumstances. On-demand has the 'ready, set, go' character of a contest rather than a realistic workplace character of 'start as soon as you can, don't leave until you are finished, let me know if you need any help, I'll come by in a while to look at your work and answer any questions, if you get stuck check with your colleagues, etc.'. Tests are often on-demand while projects, studying, writing assignments , and homework are more realistic performances. On-demand responses are drafts, not final edited works. Even the best responses will typically need some editing. When scoring, focus on what has been accomplished in the response rather than what may be omitted, or may need editing.

Using the Rubric

The rubric draws upon the *NCTM Curriculum Standards*, 1989; scorers should be familiar with these standards and have access to them during scoring. The rubric also draws upon the professionalism of teachers in two specific ways: individual experience and collegial participation in a professional consensus. Scorers act for their profession based on consensual standards.

The overarching issue for scoring is always how well the response accomplishes the prompted purpose. The 'prompted purpose' is the purpose as understood by the students. Look at the prompt from the standpoint of the student. The rubric is written to allow for variety in the ways students approach and communicate the accomplishment of the purpose. Sometimes a student will respond with an unexpected but plausible interpretation of the prompted purpose; in such cases, judge the work on its own terms consistent with the rubric. Do not mark down from a checklist of expected response features; do not mark up from a list of anticipated ingredients; rather, judge the response as a whole: how effective is it at accomplishing the prompted purpose

Within the framework of accomplishing the task's purpose, the rubric asks the scorer to appraise the response according to how well it exemplifies the mathematical power described in the first four NCTM Standards: problem solving, communication, mathematical reasoning, and connections. In addition, mathematical ideas from those described in Standards 5 through 13 for grades K-4 will be explicit and implicit in the task. Successful responses are informed by these ideas. Finally, the scorer must judge how effectively the response uses mathematical and problem solving tools and techniques as described in the Standards for grades K-4.

The rubric unfolds from the description of the 5 response. To understand the rubric, start with the 5 criterion which includes all the elements and perspectives to be evaluated in the responses. The higher and lower levels differ from the 5 level in the quality of work. While the 5 criterion is constructed with "and", the others are constructed with "or". There are many ways to exceed a 5, or to fall short, but all 5 responses accomplish the prompted purpose with high quality work.

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

6 The student's response accomplishes everything a "5" response accomplishes, but with deeper mathematical insight or eloquence, or with exceptionally well-organized reasoning, illuminating connections, powerful strategy, or other demonstration of exceptional mathematical power pertaining to the task. There are many ways a response might excel; for example, it might:

- communicate reasoning clearly for each choice in terms of constraints, or
- *find more than one combination of four kinds of fish, or*
- convincingly explain why four kinds of fish is the maximum, or
- use information from the chart effectively to further support choices, or
- incorporate special needs and variety into selection, or
- show insight into issues arising from working with several variables

5 The student's reasoning shows understanding of the two major task constraints: money and crowding, and how the selection of fish fits these constraints. The response effectively communicates the student's reasons for choosing specific fish and how the whole selection makes sense. These reasons include some reference to the additional prompted goals of variety and accounting for special needs, and the selection accomplishes these goals reasonably well.

The response reflects an understanding of combining lengths to obtain reasonable results; a student might reasonably decide to exceed the crowding constraint by a slight amount with reasons. Information from the chart is interpreted usefully; although there may be minor misinterpretations, these do not invalidate the overall selection and rationale; for example, a 5 response could include a misunderstanding of "schools," of who red tailed sharks fight with, or of whether guppies are "2/pair or 2 each." The numerical operations are used effectively to obtain reasonable results; minor computational errors can only be accepted in a 5 response if the overall results are good.

4 The student's reasoning shows an understanding of the task constraints: money and crowding. In choosing fish, the student successfully attends to at least one of these constraints. The response may lack a clear, complete explanation or demonstration of how choices were made or why they make sense; or, there is no evidence of combining lengths to evaluate crowding; or, computational errors or misinterpretation of the chart lead to results which are unreasonable; or, the response does not show at least 3 kinds of fish selected.

3 The student's reasoning shows understanding that the task has constraints of money and crowding, but in choosing fish the student fails to attend successfully to one or both of these constraints. The response communicates choices, but may explain little about how they were made or why choices make sense; or, major computational errors yield nonsensical results; or, misconceptions about the chart or the constraints invalidates the reasoning.

2 The response shows a purposeful effort to choose fish and communicate but lacks an explanation of choices; the reasoning may show understanding of one or both of the task constraints but fails to attend to either constraint. There may be little evidence of using the chart, effective understanding or use of the idea of combining lengths, or appropriate computations.

1 Little evidence of communication or attempt to address the task.

0 No response.

LEP Response includes language other than English, route to LEP scoring table

Aquarium Benchmark Responses

Computation

Within a well-reasoned response, a minor computational error (see 730339) or a misinterpretation of some aspect of the chart might not necessarily exclude the response from a "5" or "6" score. Look at the response as a whole: give greater weight to the student's mathematical understanding of the task in its entirety than to flawless computation. If a student's response exceeds the crowding or money constraint by a slight amount and the crowding or money constraint by a slight amount and the student acknowledges this discrepancy, this need not prohibit a score of "5" or "6" (see 730331).

The Chart

Some otherwise quite powerful responses reflected a misunderstanding of some element contained in the chart. The most common misinterpretations:

- meaning of "schools" (see 730258)
- do red tailed sharks fight with their own species of or other sharks?
- guppies are 2 for $3 and 2", is 2" for the pair or each?

If a students used the chart effectively except for misunderstanding one aspect of it, be flexible and evaluate the work as a whole. If it is powerful as a whole, it can score a 5.

"4" Response

- Response lacks a clear, complete explanation or demonstration of how choices were made, why they work (see 730431, 730433).
- Response contains computational errors or misinterpretation of chart (see 730338, 730871).
- Response does not show maximum variety in kinds of fish selected (see 730262, 730875).

"3" Response

- Response contains gaps, gives only minimal demonstration of how choices were made or why they work (see 730257, 730245, 730457).

Serial Number	Score	Serial Number	Score
730258	6	730245	3
730331	5	730458	3
730484	5	730457	3
730306	5	730451	3
730339	5	731068	2
731056	5	730445	2
730431	4	730256	2
730262	4	730466	2
730875	4	731001	2
730338	4	730942	1
730433	4	730244	1
730871	4	730940	1
730257	3	730993	0

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

Dear principal

I chose these fish because they get along with other fish. Also because when I imagine these fish I see buitiful ones. Thats why I chose these fish. But now I will tell you the fish I have chosen. I have chosen the guppy, the Ramire Dwarf cichlid. and the Blind cave fish. I also chose two more fishes and they are the zebra' Danio and also the marbled Hatch fish. thats the fish I chose, with the 30.00. now that I've shown you the kind of fish I bought. now then I'm going to tell you how I no these fish won't be crowded. See I've got two fish apiece of a kind and I put all the fishes I bought and put all there lenghs to gether. and it added up too 21. and the tank is 30. so thats how I know that they will not be so crowded to gether. So now that I showed you every thing I did I will end this letter by saying buy.

Sincerly

your student

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

BEST COPY AVAILABLE

May 15, 1992

Dear M.Mrs. Burk,

For the fish tank in my class, we have decided on getting 12 fish. We will get 5 Blind Cave fish because they dont take up to much space. They will only take up 15 gallons of water and ten dollars. Next we will get a pair of Ramirez Dwarf Cichlid because they get along with other fish. They will only take up four gallans of water and 10 dollars. Then last we will get five Zebra Danios because they too get along with other fish. They will take up to 6½ gallons and five dollars. So that leaves zero dollars left and 5½ gallons of water for the fish to swim around in. I think this is a good plan because the tank wont be crowed.

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only

730306

Fish	Cost	Gallons
1. Blind Cave Fish get 5	$10	15 gallons
2. Ramirez' Dwarf Cichlid get 2	$10	4 gallons
3. Zebra Danio get 5	$5	6½ gallons
	$0 left	5½ gallons left

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

730262

Dear Principal,

I have chose the fish for the 30 gallon aquarium. The first fish I bougt is a Red-Tailed Black Shark it will cost $5.00. I chose this fish because I thought this fish was very neat and I the class would enjoy watching it. This fish is 4½ inches. Next I bougit 5 Zeba Denio. This will cost $5.00 I chose this kind of fish because it is prelly and gets along with other fish. It is 1½ inches. Last I will buy 2 Ramirez' Dwarf Cichlid. This will cost $10.00. I chose this kind of fish because it is rainbow colored and gets along with other fish. It is 2 inches. It should not be to crowded in the aquaium.

Laboratory Network Program - *Alternative Assessment Toolkit* 13
Sample 4.6 *Aquarium Problem*
LNP Contributor Judy Arter NWREL

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

Laboratory Network Program—*Alternative Assessment Toolkit* 14
Sample 4.6 *Aquarium Problem*
LNP Contributor Judy Arter, NWREL

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

730457

Dale Christi
Kingman, AZ
5/14/92

Dear Dr. Fishlock,

I would like 8 zebra danios, 4 marbled hatchetfish, and 4 guppies. I chose the zebra danio because it gets along with other fish. I chose the marbled hatchetfis because they can leap 3-5 yards. I chose guppies because their very pretty. I added up how much money it would be it was $78.00 and I added up how many inches all the fish would be together they would be 24½ inches together

your friend,

① 2 of them
Red-tailed
BlAck Shark

② 2 of them
ZebrA DAnio

Dear Principal, I brought this and
I think th

Permission for use in training has been granted by New Standards Project. Materials may be copied for training purposes only.

BEST COPY AVAILABLE

Sample 4.7

Performance Assessment in Math

Greg Hall, Alberta Education, Canada

Alberta Education has developed a set of hands-on tasks to assess problem solving and communication in mathematics. Students circulate to six stations that contain hands-on materials: Seating Place (numeration, analysis, use of model); Highway to Mathematics (measurement, application); Solar Cylinder (geometry, application, analysis, synthesis); Ski Jumping (graphing, numeration, data collections, analysis, interpretation); Bucket of Beans (measurement, estimation); and Packaging (numeration, analysis, monitoring). It takes about one-and-one-half hours for students to complete all the tasks in the circuit. The same performance criteria are used for all performance tasks.

Attached is one grade 9 task and the performance criteria.

Greg Hall
"Performance Assessment in Math," 1992
Alberta Education
Box 43, 11160 Jasper Avenue
Edmonton, Alberta, Canada T5K 0L2
(403) 427-0010
FAX (403) 422-4200

Permission for use in training has been granted by Greg Hall, Alberta Education. Materials may be copied for training purposes only.

Bucket of Beans

Problem:

Estimate the number of beans in the bucket.

You should have:

large bucket of beans
magic markers
tray
small cup

Instructions:

Using any estimating strategy or population sampling techniques, estimate the number of beans in the bucket.

NOTE: Leave the station in the same arrangement it was set-up originally.

Permission for use in training has been granted by Greg Hall, Alberta Education. Materials may be copied for training purposes only.

Bucket of Beans

1. Explain the strategy you used to estimate the number of beans

2. About how many beans are there in the bucket?

DO NOT USE

	1a.	**0**				
	1b.	**1**	**2**			
	1c.	**1**	**2**			
	1d.	**1**	**2**			
	1e.	**1**	**2**			
	1f.	**1**	**2**			
	1g.	**1**	**2**			
	2.	**0**	**1**	**2**	**3**	**4**
PS	**3.**	**0**	**1**	**2**	**3**	
C	**4.**	**0**	**1**	**2**	**3**	

Permission for use in training has been granted by Greg Hall, Alberta Education. Materials may be copied for training purposes only.

MATHEMATICS PBA HOLISTIC SCORING CRITERIA

		PROBLEM SOLVING	COMMUNICATION
3	BEYOND	Analyzed and readily understood the task. Developed an efficient and workable strategy. Showed explicit evidence of carrying out the strategy. Synthesized and generalized the conclusion.	Rich, precise and clear all the time (mathematically correct, correct symbolism). Representation is very perceptive (chart, diagram, graph). Explanations are logical and appropriate.
2	AT LEVEL	Understood the task. Developed a workable strategy. Inferred (some evidence) but not always clear. Connected and applied the answer.	Appropriate most of the time, accurate, mostly clear. Representation is accurate and quite appropriate. Explanations are mostly clear and logical.
1	NOT YET AT	Partially understood the task. Appropriate strategy some of the time. Possible evidence of a plan - not clear. Partial connection of answer.	Appropriate some of the time, but may not be clear. Uses representation but not too precisely. Explanations have some clear parts.
0		Totally misunderstood. Inappropriate, unworkable strategy. No evidence of carrying out a plan. No connections of answer. Blank.	Unclear or inappropriate use of symbolism. Incorrect use of representation. Explanation is not clear. Blank.

Permission for use in training has been granted by Greg Hall, Alberta Education. Materials may be copied for training purposes only.

Sample 4.8

Science—New Directions in Assessment

California Department of Education

In conjunction with national science reform efforts, the *Science Framework for California Public Schools, Kindergarten Through Grade Twelve*, provides the foundation for systemic reform in California. To ensure that all students are given the opportunity to receive a quality science education and to achieve scientific literacy, the *Science Framework* integrates curriculum, instruction, assessment, and professional development. When writing the *Science Framework*, teachers, science educators, and scientists worked together to create a thematic and conceptual framework for science education.

The role of CLAS is to implement the recommendations of the *Science Framework* in assessment at grades 5, 8, and 10. CLAS science assessments support good curriculum and instruction and allow students the opportunity to demonstrate conceptual understanding of the big ideas of science; to use scientific tools and processes; and to apply their understanding of the big ideas to solve new problems.

In 1991, as part of large-scale field tests of performance-based tasks, students in over 2,000 schools tried out hands-on science tasks at grades 5, 8, and 11. The scenario for the attached task was the investigation of the death of the fictitious James Obechki to see if a crime had been committed. Working in groups of three, students at grade 11 analyzed an array of evidence that required them to coordinate their science abilities from biology (use a microscope to investigate properties of different hair samples) with chemistry (conduct a chromatography test to determine who wrote a note on the body) and earth science (conduct profiles and pH tests on samples of soil found on the victim's shoe). The objective of this task was to assess students' abilities to (1) use scientific processes and tools; (2) communicate thinking processes; and (3) demonstrate understanding of concepts that are connected and integrated among the sciences.

CLAS science assessments are scored by teachers who act as readers, table leaders, and chief readers. CLAS consultants, working with testing contractors and psychometricians, provide training in the scoring process for the chief readers and table leaders, who then train the readers.

Science assessment results are reported via the CLAS Science Performance Standards, benchmark descriptions against which student work can be compared. They provide a basis for teachers to make judgments about the level of accomplishment demonstrated by student work.

"Science—New Directions in Assessment, 1993"
California Department of Education
721 Capital Mall
Sacramento, CA 94244

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

CALIFORNIA LEARNING ASSESSMENT SYSTEM

1993 Performance-Based Assessment
(Originally administered in 1991 at Grade 11)

Grade 10

SCIENCE

Directions to Students: You will be completing three performance tasks. Take this answer sheet and your pencil with you as you go to each station.

A Criminal Investigation

Name:____
(Last) (First)

Sex: M or F **Date of Birth:**______________________
Month Day Year

School:____________________**District:**____________________

Partners: ____

____

© California Department of Education, Sacramento, 1993

449

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Introduction

A dead body has been found, but the cause of death is unknown. You and your partners will be working as a team of forensic scientists to examine evidence, organize data, and draw conclusions. Using the scientific equipment provided, your team will try to determine if a crime has been committed.

In your team you will complete three investigations. In order to complete the three investigations, your team will need to do the following:

- Make careful observations.
- Record and organize your data in a logical fashion.
- Make your conclusions based on your data.
- State your conclusions so that someone else will be able to understand how you arrived at your decisions.

I. The Situation

Mr. James Obechki, age 39, has been found dead on the floor of his living room. A chair has been overturned, a broken lamp is on the floor, and the rug beneath his head has red stains on it. You will be examining a variety of clues found at the death site to determine if a crime has been committed and, if so, which of the several individuals known to Mr. Obechki could have been at the site at the time of his death. Detectives at the site have gathered a set of evidence from around Mr. Obechki's body and from the living room. As a forensic scientist, you will examine this evidence and decide what might have caused Mr. Obechki's death.

GO TO NEXT PAGE

© California Department of Education, Sacramento, 1993

→ Before you begin

- **Briefly** read the descriptions of each individual in the **Personal Profile Chart**.
- You may refer to this chart at anytime during your investigations.

PERSONAL PROFILE CHART

Name	**James Obechki**	**Louise Magginity**	**Howard Jurgensen**	**Lester Simons**
Address	2204 Ocean Breeze Lane	2206 Ocean Breeze Lane	906 Sixteenth St., #31	13529 E. Ashland
Relation-ship	The deceased	Neighbor to victim, possible romantic relationship	Former business partner	Victim's cousin – owes victim $2000.00
Occupation	Owner, waterfront ocean sportfishing gear rental shop.	Librarian, public library.	Employee, Acme Cleaning Manufacturer – makes a cleaning product with borax.	Farmer.
Clothing	Wearing white cotton shirt - torn pocket, brown pants, brown argyle socks (hole in toe), no shoes.	Not known.	Ragged running shorts (grey/white), and terry cloth headband (red).	Overalls.
Pets	Goldfish, grey-striped cat.	Dog.	Miniature sheep	None
Hobby	Fishing	Golf, photography, running.	Growing plants in sand in a greenhouse and in the backyard.	Weightlifting
Vital Statistics				
Sex	Male	Female	Male	Male
Age	39	43	56	26
Height	5'10"	5'6"	6'1"	6'4"
Weight	188 lb.	135 lb.	203 lb.	209 lb.
Hair	Black – straight	Black – curly	Brown – straight	Black – straight

GO TO NEXT PAGE

© California Department of Education, Sacramento, 1993

Pre-Investigation

The District Attorney's office requested that you provide the team of detectives with a list of questions to help them organize their investigation.

1. What questions would you ask? Make sure you explain why you would ask each question.

GO TO NEXT PAGE

© California Department of Education, Sacramento, 1993

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Investigation #1 — "The Hair Bags"

The detectives working on the case went over Obechki's living room with a fine tooth comb. They collected three bags of hair samples as evidence.

Bag A	**Bag B**	**Bag C**
"Human Hair" evidence	"Unknown Hair" evidence	"Animal Hair" evidence

→ Carefully examine and compare the hair in **Bags A, B**, and **C**.

→ Design a chart in the space below to record and organize your data in a logical fashion. Put a title on your chart.

GO ON TO NEXT PAGE

© California Department of Education, Sacramento, 1993

5. What conclusions can you make from your observations of the hair samples?

STOP HERE

© California Department of Education, Sacramento, 1993

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Investigation #2 — The Note

In Mr. Obechki's hand, the detectives found a crumpled, torn piece of paper with some writing on it. Using the technique known as chromatography, you will test the ink in Mr. Obechki's pen to see if the words on the paper were written with his pen.

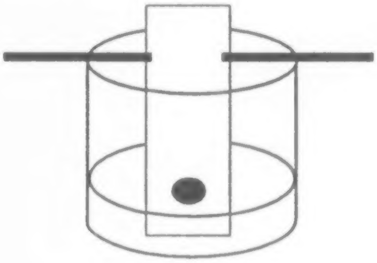
Two other pens, one belonging to Howard Jurgensen and one belonging to Louise Magginity, have been legally obtained with a search warrant. These two pens will also be tested to see if the note was written with one of them.

Chromatography is a technique used to separate the chemical parts of a solution based on the solubility of each part in a solvent which, in this case, is water.

Directions

→ Locate the following items:

- "Evidence" which has already been prepared for you. It contains a dot of ink from the paper found in Mr. Obechki's hand.
- Plastic cups in which you add water and then hang the strips of paper as you perform this investigation.
- Strips of filter paper, plastic straws, adhesive tape, and pens labeled **Obechki**, **Magginity**, and **Jurgensen**.
- Tape the top of strip "Evidence" to a straw.
- Pour a small amount of water into the cup. (See picture below.)
  Hang the strip in a cup of water. *Make sure that the dot itself does not touch the water or hang below the level of the water.*



GO ON TO NEXT PAGE

© California Department of Education, Sacramento, 1993

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

- Set up 3 more strips of filter paper with a dot of ink from each of the 3 pens (**Obechki**, **Magginity**, and **Jurgensen**) taken into evidence.
- Watch the water as it rises up the strip. Remove the strips from the water when the water almost reaches the top of the strip.
- Dry the strips on a paper towel.

→ On the chart below to show what is happening to your strips. Label the colors you see.

Chromatography Test Results Chart

Evidence	Obechki's Pen	Jurgensen's Pen	Magginity's Pen

6. Which pen, if any, matches the ink found on Mr. Obechki's note?

7. How do you know?

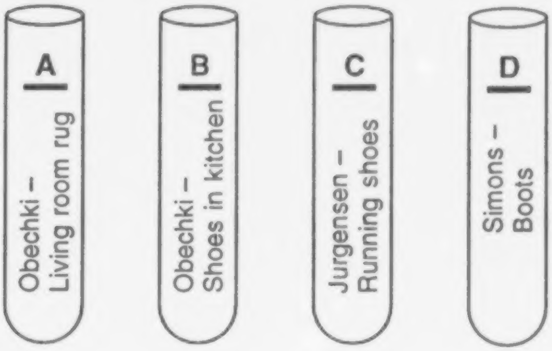
STOP HERE

© California Department of Education, Sacramento, 1993

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Investigation #3 – A Speck of Soil

Clumps of soil were found on Mr. Obechki's living room rug. Soil samples were removed from his shoes, which were found by the back door to the kitchen. Soil samples were legally obtained from Howard Jurgensen's running shoes and Lester Simons's boots on the day of Mr. Obechki's death.


A
Obechki – Living room rug
B
Obechki – Shoes in kitchen
C
Jurgensen – Running shoes
D
Simons – Boots

→ You will perform two tests to determine if any of these samples are the same.

Test 1: Soil Profile Test

When soil samples are mixed in water and shaken briefly, the soil mixture will settle to form distinct layers. Shake Vials A through D, observing how the soil in each sample settles, and record your observations in the **Soil Sample Test Chart** on the following page.

Test 2: pH Test

Use the pH test papers to determine the pH of the soil and water mixture in each vial. Use a separate paper for each test. The pH color scale is provided to assist you. Record your results in the **Soil Sample Test Chart** on the following page.

GO ON TO NEXT PAGE

© California Department of Education, Sacramento, 1993

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Soil Sample Test Chart

	Profile	pH
Vial A		
Vial B		
Vial C		
Vial D		

8. Compare the soil sample taken from the rug next to Mr. Obechki's body with the other samples of soil. What conclusion can you draw from your results?

9. What evidence supports your conclusion?

GO ON TO NEXT PAGE

© California Department of Education, Sacramento, 1993

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Putting It All Together

10. What conclusions can you make based on the evidence you have gathered? Support you answer, using the results of your investigations.

11. What additional data would you like to gather? Why?

STOP

© California Department of Education, Sacramento, 1993

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Draft Scoring Rubric

Grade 10 Performance Assessments

The following rubric was developed to assist teachers in scoring student responses to performance based assessment items. This rubric allows for a wide range of student responses and should be used as a general guideline. No one criteria should be given more weight than another.

Level 4

The student demonstrates a clear understanding of the problem and investigations performed. The student's observations are valid and demonstrate attention to detail. The quality of the data collected reflects their success at performing each experiment. The design of experiments suggested by the student are well thought out and scientifically accurate. All data is presented in an organized fashion, utilizing appropriately designed tables and charts. Answers to questions are well thought out and supported by experimental data. In general, there are no false assumptions or misleading statements made by the student. The student recognizes the need for additional testing and provides appropriate suggestions related to the problem. The conclusions reflect the student's ability to effectively analyze experimental data and draw appropriate conclusions. The student successfully proposes an explanation which clearly shows a relationship between their data and their final conclusion.

Level 3

The student demonstrates an understanding of the problem and investigations performed. The student's observations are valid and demonstrate some attention to detail. The quality of the data collected generally reflects their success at performing each task. The experiments suggested by the student are well thought out, but may have flaws in scientific design. Most data is presented in an organized fashion, utilizing appropriately designed tables and charts. Answers to most questions are well thought out and supported by experimental data. In general there are few false assumptions or misleading statements. The student recognizes the need for additional tests, but is unable to provide appropriate suggestions related to the problem. The conclusions reflect the student's ability to analyze experimental data and draw conclusions. The student successfully proposes an explanation which generally shows a relationship between their data and their final conclusion.

Level 2

The student demonstrates some understanding of the problem and investigations performed. The student's observations are vague and lack attention to detail. The quality of the data collected reflects limited success at performing the tasks. Some tasks may have been omitted. The experiments suggested by the student are unclear or riddled with flaws in scientific design. The data is recorded in a disorganized fashion, with tables and charts that are poorly designed or missing. Answers to questions are not well thought out or supported by experimental data. In general, the student makes both false assumptions and misleading statements. The student fails to recognize the need for additional tests. The student has difficulty analyzing experimental data and drawing conclusions. The student fails to propose an explanation which shows a relationship between their data and their final conclusion.

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Level 1

The student demonstrates little or no understanding of the problem or investigations performed. The student's observations are poor or missing and show no attention to detail. The quality of the data collected reflects little or no success at performing the tasks. The experiments suggested by the student were difficult to follow, or missing. The data collected is haphazardly recorded, or missing. Answers to questions are implausible and not related to the experimental data. In general, the student makes many false assumptions and misleading statements. The student fails to recognize the need for additional tests. The student is unable to analyze experimental data or draw conclusions. The final conclusion is unrelated to the tasks performed, or missing.

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Sample 4.9

Mapping the Blue Part

Kentucky Department of Education

The attached task and related scoring guide is from the 1993 grade 8 assessment. It was one of three tasks that required some group and some individual work.

Performance Events, 1992-93
Kentucky Department of Education
Capital Plaza Tower
500 Mero St.
Frankfort, KY 40604
(502) 564-4394

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

PERFORMANCE ASSESSMENT STUDENT INSTRUCTION/RESPONSE FORM

TASK: M/S 1 "Mapping The Blue Part" GRADE 8

STUDENT NAME:____

SCHOOL NAME:____

GENERAL INSTRUCTIONS:

Your task is to determine the location of specific features of the simulated ocean floor. The features include the continental shelf, continental slope, an ocean trench, and a mid-oceanic ridge.

You may use up to 20 minutes to complete the "Group Instructions" given below.

MATERIALS:

- black box with simulated ocean floor and scale
- height finder with sliding ring

GROUP INSTRUCTIONS:

You will have approximately 20 minutes to complete steps 1 through 3 as a group.

1. Begin height readings by inserting the height finder at the coordinate (H-1). To determine the height, measure the distance (in centimeters) from the BOTTOM of the sliding ring to the TOP of the height finder by using the scale on the right side of the box. Each member of the group should record the measured height on his/her own grid on page 4.

2. Continue taking and recording height readings for all of the points in column "H".

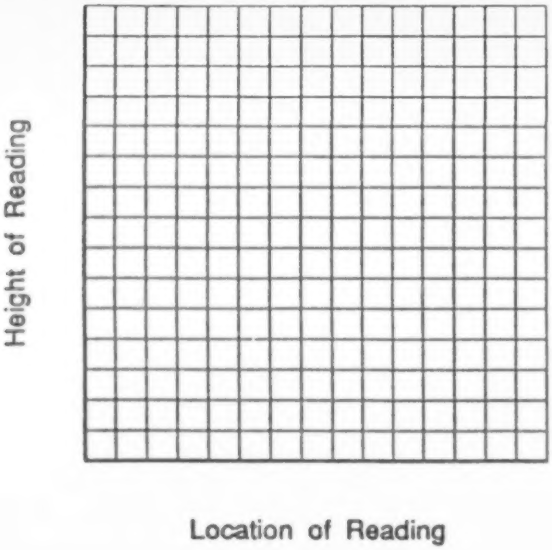
3. When you have completed column "H", you are to take and record height readings for all of the points in row "12".

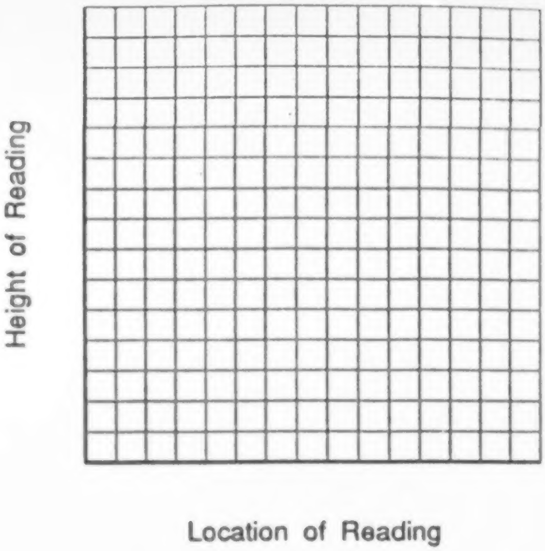
After you finish steps 1 through 3, you will open this booklet and answer questions 4 - 7 on your own.

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

You will have approximately 25 minutes to complete questions 4 - 7 on your own.

4. Using the data you have collected, produce graphs which represent a profile of the ocean floor.





5. Think about the features of the ocean floor: continental shelf, continental slope, ocean trench, and mid-oceanic ridge. Using your readings, make a "map" of the ocean floor, identifying the possible locations of ocean floor features. Use the rectangle below for your map. Label the cordinates along the edges to show the location of these features. What additional readings would you take to verify your inferences (ideas)? Why would you take these readings?

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

6. What are some real life applications of mapping the ocean floor? What benefit could these applications provide?

7. If you were an oceanographer wanting information concerning the ocean floor would you use this method? Why or why not? What additional investigations would you use to map the ocean floor?

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

SCORING GUIDE
KIRIS Performance Event, Grade 8
1992-93
Task: M/S1 - Mapping the Blue Part

Group Activity

	ACTUAL	REVERSE
H1	8.5	3.0
H2	8.0	3.5
H3	7.5	4.0
H4	7.5	4.0
H5	7.5	4.0
H6	6.5	5.0
H7	6.0	4.5
H8	6.0	4.5
H9	6.5	5.0
H10	6.0	4.5
H11	6.5	5.0
H12	6.5	5.0

	ACTUAL	REVERSE
A12	4.5	7.0
B12	4.5	7.0
C12	7.0	4.5
D12	6.5	5.0
E12	6.5	5.0
F12	6.5	5.0
G12	6.5	5.0
H12	6.5	5.0
I12	6.5	5.0
J12	6.5	5.0

SCORE	DESCRIPTION
4	All points are correct (see table) ± .50cm.
3	13-20 points are correct.
2	5-12 points are correct.
1	0-4 points are correct.
0	BLANK

NOTE: Reverse readings result in loss of 1 point, but not for scores of 2 or 1.

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

KENTUCKY HOLISTIC SCORING RUBRIC

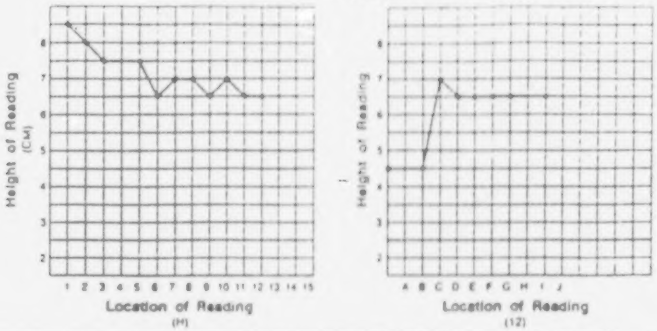
PERFORMANCE EVENTS

CATEGORY 4	• The student completes all important components of the task and communicates ideas clearly. • The student demonstrates in-depth understanding of the relevant concepts and/or processes. • Where appropriate, the student chooses more efficient and/or sophisticated processes. • Where appropriate, the student offers insightful interpretations or extensions (generalizations, applications, analogies).
CATEGORY 3	• The student completes most important components of the task and communicates clearly. • The student demonstrates understanding of major concepts even though he/she overlooks or misunderstands some less important ideas or details.
CATEGORY 2	• The student completes some important components of the task and communicates those clearly. • The student demonstrates that there are gaps in his/her conceptual understanding.
CATEGORY 1	• The student shows minimal understanding. • The student is unable to generate strategy, or answer may display only recall effect. Answer lacks clear communication. • Answer may be totally incorrect or irrelevant.
CATEGORY O	• Blank/no response.

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

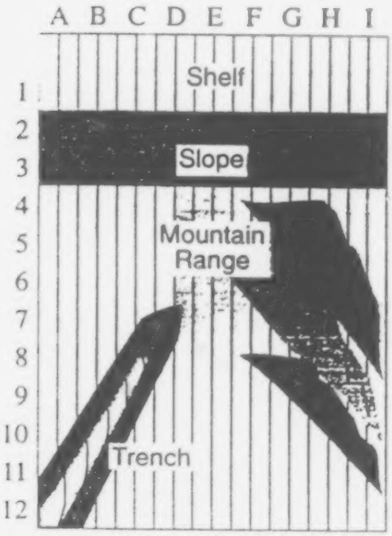
Question 4

Individual Activity



SCORE	DESCRIPTION
4	Student's graph is an accurate representation of group data, with both axes reasonably divided.
3	Student's graph is a nearly accurate representation of group data.
2	Student graphs some points accurately; axes may not be labeled correctly.
1	Student shows minimal understanding of graphing.
0	BLANK

Question 5



SCORE	DESCRIPTION
4	Student maps ocean floor, representing the major features, and explains why and where he/she would take additional readings.
3	Student maps ocean floor, representing most major features, and explains why and where he/she would take additional readings (may exclude one feature).
2	Student maps ocean floor, representing some major features, and may or may not explain why and where he/she would take additional readings (may exclude two features).
1	Student recognizes location of only one or two features, and does not explain additional readings.
0	BLANK

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

Question 6

SCORE	DESCRIPTION
4	Student recognizes more than one application and related potential benefits.
3	Student recognizes one application and its potential benefit(s).
2	Student recognizes application, without naming benefits.
1	Student fails to recognize any application.
0	BLANK

Question 7

SCORE	DESCRIPTION
4	Student recognizes method's limitations and chooses another relevant method to map the ocean floor.
3	Student recognizes method's limitations but is unable to develop another method to map the ocean floor.
2	Student believes method used has no limitations.
1	(Does not apply.)
0	BLANK

Permission for use in training has been granted by the Kentucky Department of Education. Materials may be copied for training purposes only.

Sample 4.10

Writing in Chemistry

CRESST

The attached chemistry assessment is one of two examples developed by CRESST to illustrate an assessment approach that includes prior knowledge, primary source materials, and writing to promote thinking. (The other example is U.S. history.)

Attached are the writing assignment and an overview of the scoring procedure. The reading passage is not included.

Eva Baker, et. al.
"CRESST Performance Assessment Models," 1992
National Center for Research on Evaluation, Standards, and Student Testing
Center for the Study of Evaluation
UCLA Graduate School of Education
145 Moore Hall
Los Angeles, CA 90024
(310) 206-1532

Permission for use in training has been granted by the National Center for Research on Evaluation, Standards, and Student Testing. Materials may be copied for training purposes only.

Example 3.6 Chemistry Writing Prompt

Name____________

Writing Assignment

Imagine you are taking a chemistry class with a teacher who has just given the demonstration of chemical analysis you read about earlier.

Since the start of the year, your class has been studying the principles and procedures used in chemical analysis. One of your friends has missed several weeks of class because of illness and is worried about a major exam in chemistry that will be given in two weeks. This friend asks you to explain everything that she will need to know for the exam.

Write an essay in which you explain the most important ideas and principles that your friend should understand. In your essay you should include general concepts and specific facts you know about chemistry, and especially what you know about chemical analysis or identifying unknown substances. You should also explain how the teacher's demonstration illustrates important principles of chemistry.

Be sure to show the relationships among the ideas, facts, and procedures you know.

Permission for use in training has been granted by the National Center for Research on Evaluation, Standards, and Student Testing. Materials may be copied for training purposes only.

Contextual prompts to heighten student understanding of the task

Our research suggests that the essay prompt should be highly contextualized to help students reveal their understanding of the issues they are to address. We tried various kinds of prompts, some with a strong narrative context, for example, "You are a farmer in Illinois in the 1850's..." as well as those which ask students to write a specific explanation to another person, such as "You want to explain to your cousin...." Although we found no significant improvement from using such prompts, they may enhance student fluency. We recommend contextual prompts to heighten student understanding of the task and to provide an audience (even though we understand that school-based writing has only limited "real life" relevance) as this is becoming common practice in literature-based writing programs.

The CRESST essay scoring rubric. We employ a criterion-referenced scoring strategy. Papers are judged in terms of pre-specified standards rather than in relationship to one another.

Scoring rubric for the essay

The scoring rubric for the essay consists of six scales or dimensions:

1. General Impression of Content Quality;
2. Prior Knowledge;
3. Number of Principles or Concepts;
4. Argumentation;
5. Text; and
6. Misconceptions.

General Impression of Content Quality rates the overall competence the student exhibits in answering the question. This dimension focuses on the way students demonstrate their understanding of the concepts (and in history, the historical era discussed).

Prior Knowledge is a judgment about the amount of relevant information that is not explicitly in the primary text material, which the student uses to illustrate or support his or her perspective.

Permission for use in training has been granted by the National Center for Research on Evaluation, Standards, and Student Testing. Materials may be copied for training purposes only

Scoring rubric for the essay

Number of Principles or Concepts is a count of the major principles or concepts that students use with comprehension in their essay.

Argumentation is a measure of how well the student organizes knowledge to present a convincing argument or a coherent explanation.

Text (proportion of essay using text-based detail) is an assessment of how much information presented in the essay is derived from primary text materials.

Misconceptions asks raters to judge the extent to which students reveal major misunderstandings of concepts or of combinations of facts.

A complete guide to the essay scoring rubric is in Chapter 6 of this handbook.

Scales of the scoring rubric are differentially useful

Each essay is scored on all six dimensions using a scale ranging from 0 to 5. Our studies indicate that, in general, the best essays receive low-to-moderate scores for *Text (*proportion of essay using text based detail), with high scores on the other scales. Note that a high score on *Misconceptions* indicates few or no misconceptions.

The scales of the scoring rubric are differentially useful. They may be weighted differently depending upon the particular objectives of the assessment. For example, if reading comprehension or learning from new materials is emphasized, then there may be a reason to look for high scores on the *Text* dimension. When modifying the scoring rubric, directions to students should be revised to match instructional goals.

High levels of interrater reliability

We have achieved high levels of interrater reliability for the scoring rubric across different topics and different raters (see Chapter 3). The lowest reliability estimates have been associated with the dimension of *Misconceptions* (correlation of .68).

Permission for use in training has been granted by the National Center for Research on Evaluation, Standards, and Student Testing. Materials may be copied for training purposes only.

BEST COPY AVAILABLE

Sample 5.1

ASSESSMENT AND TECHNOLOGY VIDEOTAPE

BANK ST. COLLEGE OF EDUCATION

The following criteria were used with a grade 11 class at Brooklyn Technical High School in New York. To explore important concepts in applied physics, students worked in teams to design and construct motorized devices that produce at least two simultaneous motions in different directions to accomplish an action. When the projects were finished, students were required to demonstrate and explain how their devices work and why they made certain design decisions, using their understanding of physics concepts. Students were also required to give their devices a test run to demonstrate how well their devices worked and to explain what needed to be modified to make them work better.

Sample student performances are available on a videotape to practice scoring.

Dorothy Bennet
Assessment and Technology Videotape
Center for Technology in Education
Bank St. College of Education
610 W. 112th St.
New York, NY 10025

Permission for use in training has been granted by Dorothy Bennet. Materials may be copied for training purposes only.

Refined Criteria: July 19, 1993 1

SCORING CRITERIA

Thinking Skills

UNDERSTANDING

☐ Understands relationships between variables and topics covered.

☐ Locates pertinent information to solve problems. Makes observations about the work.

☐ Explains known principles/concepts/theories and how they fit into work: -uses examples to demonstrate knowledge -uses relevant terminology -applies formulas accurately and appropriately -uses diagrams, and graphs appropriately.

CRITICAL THINKING/ INQUIRY:

☐ Identifies the problem.

☐ Justifies decisions made (thinking process evident).

☐ Makes analysis/inferences/predictions based on the work.

☐ Responds appropriately to unanticipated problems.

☐ Asks informed questions about the work/curiosity.

☐ Uses resources creatively.

☐ Has innovative ideas.

Communication/ Presentation Skills

CLARITY AND COHERENCE OF PRESENTATION:

☐ Uses visual aids to make explanations clear.

☐ Explains calculations and reviews components of final product.

☐ Uses clear and concise language.

☐ Organizes materials systematically.

☐ Uses conventions that make student's train of thought evident.

PRESENTATION AESTHETICS:

Written Work

☐ Work is legible and neat; shows pride in work.

☐ Pays attention to details.

☐ Grammar and spelling.

Presentations and Oral Work

☐ Body language and poise—addresses audience/eye contact.

☐ Is confident with/has ownership of material.

☐ Has rehearsed and/or smooth delivery.

☐ Craftsmanship—Puts components of product together well.

☐ Uses appropriate medium or materials to convey ideas.

☐ Pays attention to details.

Work Management/ Interpersonal Skills

TEAMWORK (FOR GROUP WORK ONLY):

☐ Group members show interest in project/evidence that all members played a part in work.

☐ Labor is divided equally, or an obvious attempt to do so is made.

THOROUGHNESS/EFFORT:

☐ Completes work and meets requirements of task.

☐ Shows evidence of hard work in quality of final product or in explanations about work.

☐ Adds aspects that go beyond what is required in original assignment.

☐ Plans work, reports on activities.

REFLECTIVENESS:

☐ Includes individual or group opinion about process (steps involved in completing work)

☐ Personal Voice: Explains the origins of her/his ideas—indicates originality and ownership of work.

☐ Reflects on interpersonal relations within the group (when working in groups).

☐ Reflects on own limitations and strengths as learners.

☐ Points out significance of work for own learning.

☐ Puts work in a larger context or relates work to real-life.

Permission for use in training has been granted by Dorothy Bennet. Materials may be copied for training purposes only.

STEPS FOR SCORING WORK

Familiarize yourself with the criteria	Connect an idea with each of the eight subcatagories.
Read through or view the work	Get a feeling for the assignment and the student.
Look for examples of the criteria in the work	Go over the work carefully and pick out specific episodes that exemplify a criterion. Note if the episode is positive or negative (Valence, as explained below).
Give the piece of work a score for each of the criteria	Base the score on the examples you found in the work. Use the guidelines below for giving scores.
Write a rationale for each of the scores you gave	Use the examples you found in the work to write a few sentences of explanation of why you gave the score you did.
Discuss your scores and rationales with another person	Make sure you are seeing the things that another teacher is seeing and that you have noticed everything.
Come to a consensus on scores with the other person	Use the chance to see through someone elses eyes to make your scores more reliable.

GIVING NUMERICAL SCORES

VALENCE RATINGS ARE ++, +, -, --
++: Episode was a Excellent Example of the Criteria
+: Episode was a Good Example of the Criteria
-: Student Missed an Opportunity to show the Criteria
--: Episode was Counterproductive to showing the Criteria

SCORES RANGE FROM 0 TO 5
0: Not Applicable
1: Poor; Mostly Counterproductive (--s), also many Missed Opportunities (-s)
2: Needs More Work; Mostly Missed Opportunities (-s)
3: Acceptable; Few Excellent (++s) or Counterproductive (--s) examples, Balance between Good Examples (+s) and Missed Opportunities (-s)
4: Good; Mostly Good Examples (+s), but allows some Missed Opportunities (-s)
5: Superb; Mostly Excellent (++s) or Good (+s) Examples, few or no Missed Opportunities (-s)

Permission for use in training has been granted by Dorothy Bennet. Materials may be copied for training purposes only.

EXAMPLES OF CRITERIA

Presentation Piece

Scorer:______________________ Project:______________________

Valence: Wonderful Example: ++ Good Example: + Missed Opportunity: - Counterproductive: --

THINKING SKILLS U-Understanding CT-Critical Thinking	**COMMUNICATION/ PRESENTATION SKILLS** Cl-Clarity and Coherence P-Presentation Aesthetics	**WORK MANAGEMENT/ INTERPERSONAL SKILLS** TW-Teamwork Th-Thoroughness/Effort R-Reflectiveness

Permission for use in training has been granted by Dorothy Bennet. Materials may be copied for training purposes only.

Sample 5.2

Assessment of Laboratory Skills in High School Science

Rodney Doran, State University of New York at Buffalo

The author has developed laboratory performance assessments in chemistry, physics, and biology. These tasks were constructed as part of an NSF-supported project in collaboration with Dr. Darrell Bock of the University of Chicago. Six tasks are presented in each content area. Each task has two parts that take a total of 80 minutes. In Part A, students are given a problem to solve and are directed to state an appropriate hypothesis, develop a procedure for gathering relevant observations or data, and propose a method for organizing the information collected. These plans are collected prior to beginning Part B.

In Part B, students are given predeveloped procedures to collect information on the same problem as in Part A. This procedure ensures that success on Part B is not dependent on Part A. In Part B, student work is examined for quality of observations or data, performance of appropriate calculations, and formulation of valid and consistent conclusions from experimental results.

The same criteria are used for all tasks.

Attached is one of the chemistry tasks, plus the performance criteria.

Dr. Rodney Doran
Graduate School of Education
State University of New York at Buffalo
Buffalo, NY 14260
(716) 645-3171

Permission for use in training has been granted by Rodney Doran. Materials may be copied for training purposes only.

MELTING POINT

Identification of pure substances

Introduction

This laboratory test presents a problem. Your task in Part A is to plan and design an experiment to solve the problem. You will have **30 minutes** to complete Part A. At the end of the 30 minutes, your answer sheet will be collected. You will then receive separate directions for Part B. In Part B you will use materials and equipment provided in the laboratory kit to collect experimental data for this problem. You may wish to do your preliminary planning on the sheet labelled "Working Copy." Write your final plan on the appropriate answer sheet in your test booklet.

Problem

Chemists often use the physical properties of substances, e.g. melting point or boiling point, to identify substances. Your task is to plan and design an experiment, using the materials listed below and the property of melting point, to identify two substances (samples A and B).

a) Under the heading PROCEDURE list in order the steps of the procedure you will use to solve the problem. You may include a diagram to help illustrate your plans for the experiment. Include any safety procedures you would follow.

b) Construct a DATA TABLE or indicate any other method that you could use to record the observations and results that will be obtained.

PLEASE NOTE: In Part A you are NOT to proceed with any part of the actual experiment. You are just to plan and organize a way to investigate the problem.

Materials

- rubber bands
- mortar and pestle
- spatula
- support stand
- 250 ml beaker
- boiling chips
- samples labelled A and B
- capillary/ melting-point tubes
- thermometer
- gauze mat
- clamps
- bunsen burner
- one-holed cork/rubber stopper

Permission for use in training has been granted by Rodney Doran. Materials may be copied for training purposes only.

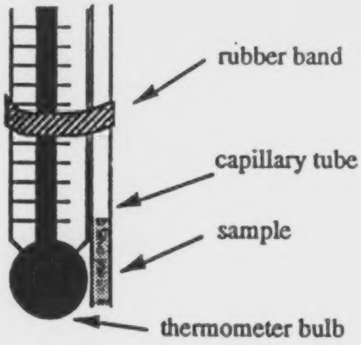
BEST COPY AVAILABLE

MELTING POINT

Identification of pure substances

Procedure

1. Use a mortar and pestle to grind a small sample of material labelled A.
2. Fill a capillary tube with the sample to a depth of 5 mm. This can be accomplished by the following steps:
    - (a) Insert the tube into the sample
    - (b) Invert the tube
    - (c) Tap the tube gently
3. Attach the capillary tube to the bulb of the thermometer using the rubber band (see diagram below).
4. Set up the apparatus as shown in the diagram on the previous page.
5. Heat the water gently and observe the temperature at which the substance melts. Record this data.
6. Repeat the procedure for sample B. Record your results.
7. Raise your hand to contact your instructor. The instructor will check to see that you have completed steps 1 through 6, then give you a sheet labelled "Melting Points of Substances." Using your data and the information in the "Melting Points of Substances" sheet, determine which substance was sample A and which was sample B.



Magnified version of melting point set-up

Safety

(1) Hot objects must be handled with tongs.

(2) Be careful handling the capillary tube. It is very fragile.

Permission for use in training has been granted by Rodney Doran. Materials may be copied for training purposes only.

Figure 1: Scoring Form for Science Laboratory Test

School/Student ID No. __________ Reader ID No. __________ Date __________ Time __________

Subject Area B C P Task 1 2 3 4 5 6

1. Please circle the NA code if a skill is not assessed in a particular area.
2. The NR code is to be circled when no attempt to respond to the question is apparent.
3. You may check each element present (✓) and sum up to determine a student's score for each skill.
4. There is no need to determine a total score for a student.

Part A: Experiment Design

1. STATEMENT OF HYPOTHESIS NR 0 1 2 3 4 5 NA

- Effect linked to variable ____
- Directionality of effect ____
- Expected effect/change ____
- Independent variable ____
- Dependent variable ____

2. PROCEDURE FOR INVESTIGATION NR 0 1 2 3 4 5 NA

- Resolves experimental problem/feasible ____
- Sequenced and detailed plan ____
- General strategy ____
- Safety procedures ____
- Use of equipment/diagram of set-up ____

3. PLAN TO RECORD AND ORGANIZE OBSERVATIONS/DATA NR 0 1 2 3 4 5 NA

- Space for measured/calculated data ____
- Matched to plan ____
- Organized sequentailly ____
- Labelled fully (units included) ____
- Variables identified ____

Part B: Experiment Report

4. QUALITY OF OBSERVATIONS/DATA NR 0 1 2 3 4 5 NA

- Consistent data ____
- Accurate measurements/observations ____
- Completed data table ____
- Correct units ____
- Qualitative description ____

5. GRAPH NR 0 1 2 3 4 5 NA

- Curve is appropriate to data trend ____
- Points plotted accurately ____
- Appropriate scale (units included) ____
- Axes labelled with correct variables ____
- Has an appropriate title ____

6. CALCULATIONS NR 0 1 2 3 4 5 NA

- Calculated accurately ____
- Substituted correctly into relationship ____
- Relationship stated or implied ____
- Units used correctly ____
- Used all data available ____

7. FORMS A CONCLUSION FROM THE EXPERIMENT NR 0 1 2 3 4 5 NA

- Consistent with scientific principle ____
- Sources of error ____
- Consistent with data ____
- Relationship among variables stated ____
- Variables stated in conclusion ____

Rodney Doran, SUNY@Buffalo

Permission for use in training has been granted by Rodney Doran. Materials may be copied for training purposes only.

BEST COPY AVAILABLE

Teacher Information

CHEMISTRY TASK 5

MELTING POINT

Identification of pure substances

MATERIALS:

rubber bands	capillary tubes (melting point tubes)
mortar and pestle	thermometer
support stand	clamp
250 ml beaker	bunsen burner
boiling chips	gauze mat
samples labelled A, and B	spatula

NOTES:

(1) Use the following substances:

A :	lauric acid	44 degrees C
B :	napthalene	80 degrees C

(2) Place approximately 3 grams of each substance in labelled beakers.

(3) Use great care when inserting the thermometer into the stopper (glycerine may help).

(4) Make sure the stopper does not block the melting-point temperature for the above substances.

Permission for use in training has been granted by Rodney Doran. Materials may be copied for training purposes only.

Sample 5.3

Science Portfolio (GSE)

California Department of Education

The Golden State Exam (GSE) science portfolio is a collection of student work produced during a year of high school biology, chemistry, or second-year coordinated science. It allows students to present for evaluation a broader representation of performance exhibiting depth of conceptual and procedural knowledge. This is an optional component for the GSE. Scores are combined with the multiple-choice, short-answer, open-ended, and laboratory performance portions of the GSE only if it would serve to improve the student's overall score.

Attached are several pages from the Teachers Handbook which discuss what should go in the portfolio, examples of entries, one sample self-reflection page, and performance criteria.

"Golden State Examination Science Portfolio—A Guide for Teachers," 1994
California Department of Education
721 Capital Mall
Sacramento, CA 94244

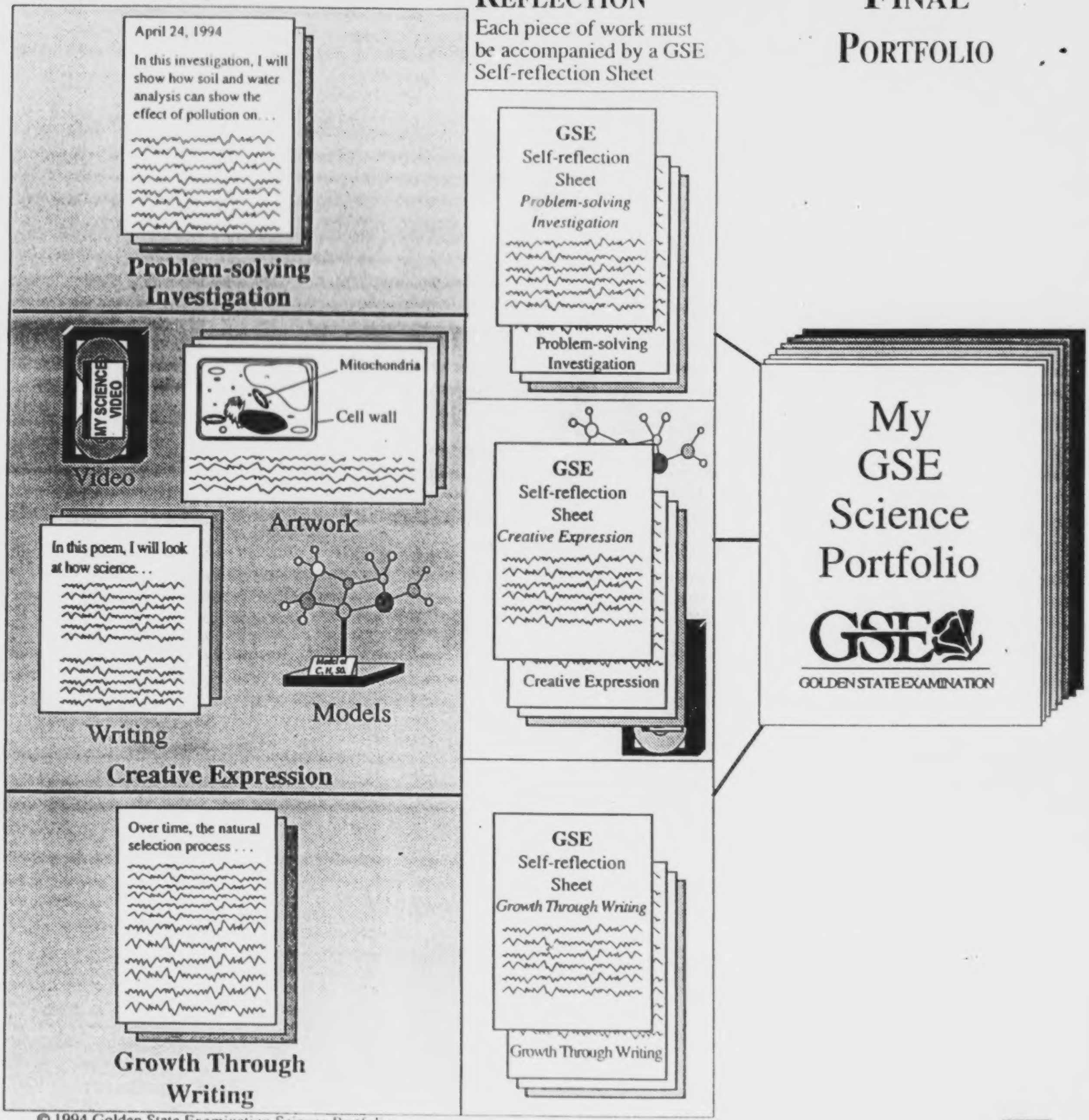
Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

ENTRY OVERVIEW

ENTRY
Work must fulfill the appropriate entry requirements

SELF-REFLECTION
Each piece of work must be accompanied by a GSE Self-reflection Sheet

SUBMIT FINAL PORTFOLIO


April 24, 1994
In this investigation, I will show how soil and water analysis can show the effect of pollution on . . .
Problem-solving Investigation
MY SCIENCE VIDEO
Video
Mitochondria
Cell wall
Artwork
In this poem, I will look at how science . . .
Writing
Models
Creative Expression
Over time, the natural selection process . . .
Growth Through Writing
GSE Self-reflection Sheet Problem-solving Investigation
Problem-solving Investigation
GSE Self-reflection Sheet Creative Expression
Creative Expression
GSE Self-reflection Sheet Growth Through Writing
Growth Through Writing
My GSE Science Portfolio
GSE
GOLDEN STATE EXAMINATION

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

BEST COPY AVAILABLE

SAMPLE ACTIVITIES

The following sample activities are not intended to be comprehensive or exclusive. Rather, they are examples that illustrate a wide variety of engaging activities modeling the exemplary instruction in science classrooms in California.

ENTRIES

	Student Generated Laboratory Investigation	Field Experience	Research Investigation
Problem Solving Investigation	Students are asked to determine which household cleaning fluids control the growth of bacteria most effectively. The teacher lists the material he or she will provide, i.e., agar plates. Students design the experimental procedures, data collection and display techniques; carry out the procedure and make meaningful applications to related issues. *Contributed by Kasey Smith-Penner, Sequoia High, Redwood City*	Students are asked to conduct research on a 1 meter plot of land in the local neighborhood. They determine what aspects of the soil, flora, fauna, and climate to study; how to make measurements and collect information; how to display data; and what to conclude from the study. Finally, they are asked to make comparisons between their findings and relevant environmental issues in their communities. *Contributed by Joe Mahood, Aragon High School, San Mateo*	Students make predictions about the risk of developing a particular ailment (they may choose from a list, i.e., cancer AIDS, heart disease, lead poisoning, etc.) in their community. They must research the disease, compare local to national risk factors, and develop an information pamphlet for their communities. *Adapted from an activity by Netta Freeman, Paduca High School, Tilghman, Kentucky*
	Students are asked to separate and collect the contents of a mixture of soil. As above, the teacher provides a list of materials he or she will provide, all other aspects of the investigation, analysis, conclusions, and relevant applications are up to the student and his or her partners. *Contributed by Steven Unterholzner, Encina High School, Sacramento and Jayson Chang, Mt. Diablo High School, Concord*		
	Games	Video	Art & Poetry
Creative Expression	Students produce games highlighting the significance of periodic properties. In one example, the object of the game is to reach the nucleus of an atom by answering questions that relate periodic properties to the real world, i.e., explanation of everyday natural phenomenon using atomic structure and characteristics. A written explanation accompanies the game. *Contributed by Steve Harness, Kingsburg High School, Kingsburg*	Students present a socially relevant topic in an easy to understand visual format that "gets the message across." A student produces a "clay-mation video" depicting the infection, replication, and damage of the AIDS virus in a well explained and detailed step by step presentation. *Contributed by Kasey Smith-Penner, Sequoia High, Redwood City.*	Students produce a drawing or painting of an environmental issue or sequence of events and write a haiku explaining their view of its relevance to them personally. To accompany the creative pieces, the student writes an explanation of the issue depicted in the art and poetry, highlighting the significant scientific concepts. *Contributed by Marian Gonzalez, Lowell High, SF*
	Progressive Writing	Current Events	Original Stories
Growth Through Writing	Students write monthly essay responses focussing on a particular concept or theme of science, i.e., energy, evolution, or ecological relationships. The questions may vary or repeat, but must be revised to include newly gained knowledge as the year progresses. Finally, the student summarizes his or her learnings in a comprehensive writing about the concept. *Contributed by Patty Kreikemeier, Santana High School, San Diego.*	Students summarize newspaper and magazine articles throughout the school year, focusing on a particular issue or scientific concept. They continually revise their writings and tie them together in a culminating essay. *Contributed by Terry Shandhan, Milliken High School, Long Beach*	Students write a series of short stories or "quick writes" explaining the significant effects of the properties of water. The writings are revised to reflect how the properties effect various environmental factors. Finally, the student describes life on an imaginary planet with certain unique features, and how the properties of water effect the flora, fauna, topography, and climate there. *Contributed by Dick Filson, - High School, Stockton*

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

GSE SELF-REFLECTION SHEET

Entry: Creative Expression

1. Identify the scientific concept you are presenting in your creative selection. (Example: genetic diversity or conservation of matter.)

2. Describe why this type of expression (model, written entry, video, poem etc.) is an excellent way to show the scientific idea you are presenting.

3. Describe how this entry shows the scientific concept. If appropriate, also describe the relationship between this concept and other scientific concepts.

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

STANDARDS OF ACHIEVEMENT

The results of the GSE are reported in terms of statewide standards which identify six levels of performance ranging from minimal to high honors. These standards establish benchmarks of the quality of performance against which student work can be compared. The following Standards of Achievement provide a concrete set of challenging performance goals which allow teachers and students to identify strengths and deficiencies and to use these comparisons to foster scientific understanding.

Standards of Achievement

Level 6 Student work at this level shows insightful reflection and mastery of scientific ideas and principles. There is exceptional evidence that entries have been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. The problem-solving investigation shows excellent student-generated experimental design and accurate data collection. Qualitative and quantitative analyses and explanations of conclusions show a high level of reasoning. Evidence of productive collaboration show improved understanding of scientific ideas. Student work shows significant applications drawn from outside experiences and previously learned science to relate scientific ideas to real-world situations. The student's creative expression clearly shows excellent understanding of a scientific idea that is enhanced by the presentation. Written work shows excellent growth and understanding of scientific ideas through comparison of original and revised samples of work. All aspects of the tasks are complete and contribute to the overall excellence of the portfolio. Student work displays careful attention to detail and excellent communication skills.

Level 5 Student work at this level shows thoughtful reflection and strong understanding of scientific ideas and principles. There is convincing evidence that entries have been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. The problem-solving investigation shows very strong student-generated experimental design and accurate data collection. Qualitative and quantitative analyses and explanations of conclusions show strong reasoning. Evidence of productive collaboration shows improved understanding of scientific ideas. Student work shows relevant applications drawn from outside experiences and previously learned science to relate a scientific idea to real-world situations. The student's creative expression clearly shows strong understanding of scientific ideas enhanced by the presentation. Written work shows convincing growth in understanding scientific ideas through comparison of original and revised samples of work. All aspects of the tasks are complete and contribute to the overall excellence of the portfolio. Student work displays careful attention to detail and commendable communication skills.

Level 4 Student work at this level shows relevant reflection and good understanding of scientific ideas and principles. There is ample evidence that entries have been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. The problem-solving investigation shows strength in student-generated experimental design and appropriate data collection. Qualitative and quantitative analyses and explanations of conclusions show a moderately high level of reasoning. Evidence of adequate collaboration shows some improved understanding of scientific ideas. Student work shows valid applications drawn from outside experiences and previously learned science to relate a scientific idea to real-world situations. The student's creative expression may show good understanding of scientific ideas that may be enhanced by the presentation. Written work shows some growth in understanding of a scientific idea through comparison of original and revised samples of work. All major aspects of the tasks are complete and contribute to the overall quality of the portfolio. Student work displays some attention to detail and solid communication skills.

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Level 3 Student work at this level shows some reflection and basic understanding of scientific ideas and principles. There is some evidence that entries have been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. The problem-solving investigation shows basic skills in student-generated experimental design and data collection. Qualitative and quantitative analyses and explanations of conclusions are incompletely substantiated. Evidence of collaboration shows some improved understanding of scientific ideas. Student work shows some applications drawn from outside experiences and previously learned science to relate a scientific idea to real-world situations. The student's creative expression shows basic understanding of a scientific idea but may not be enhanced by the presentation. Written work attempts to show some growth in understanding of scientific ideas through comparison of original and later samples. Some aspects of the tasks are complete. Student work shows little attention to detail and basic communication skills.

Level 2 Student work at this level shows limited reflection and limited understanding of scientific ideas and principles. There is little evidence that entries have been revisited or improved to reflect newly-gained knowledge, skills, and increased depth of understanding. The problem-solving investigation may not show skills in student-generated experimental design and data collection. Qualitative and quantitative analyses and explanations of conclusions, if present, are flawed or incomplete. Evidence of collaboration, if present, shows limited improvement in understanding of scientific ideas. Student work shows few applications drawn from outside experiences and previously learned science to relate a scientific idea to real-world situations. The student's creative expression may show little understanding of a scientific idea or may not be enhanced by the presentation. Written work shows limited growth in understanding of scientific ideas through comparison of original and later samples. Tasks are largely incomplete. Student work lacks attention to detail and shows limited communication skills.

Level 1 Student work at this level lacks meaningful reflection and shows minimal understanding of scientific ideas and principles. There may not be evidence that entries have been revisited and entries may not reflect newly-gained knowledge, skills, or increased depth of understanding. The problem-solving investigation may not show skills in student-generated experimental design and data collection. Qualitative and quantitative analyses may not be present or are largely incomplete and explanations of conclusions are not supported or substantiated. Little or no evidence of meaningful collaboration is present. Student work shows minimal applications to real-world situations. The student's work may show little or no creativity or may show minimal understanding of a scientific idea. Written work shows little or no growth in understanding of a scientific idea through comparison of original and later samples. Tasks are mostly incomplete. Student work lacks attention to detail and shows ineffective communication skills.

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

ENTRY SCORING GUIDE

Problem-solving Investigation

Level 6 Student work and self-reflection show the student's exemplary skill in experimental design. Excellent observations and data analysis indicate extensive knowledge of the scientific ideas presented. Analyses and conclusions are supported by observations and data and show a high level of reasoning. The entry includes conclusive evidence that the student worked cooperatively with others. The self-reflection sheet clearly identifies how working with others improved the student's understanding of the scientific ideas presented. The student's investigation makes insightful applications to real-world situations. All aspects of the task and analysis are completed thoroughly. Written expression is exemplary and contributes to clear, coherent and communication.

Level 5 Student work and self-reflection show the student's strong skill in experimental design. Observations and data analysis indicate very good knowledge of the scientific ideas presented. Analyses and conclusions are supported by observations and data and show a high level of reasoning. The entry includes strong evidence that the student worked cooperatively with others. The self-reflection sheet clearly identifies how working with others improved the student's understanding of the scientific ideas presented. The student makes generally valid applications to real-world situations. All aspects of the task and analysis are complete. Written expression is very good and contributes to clear and coherent communication.

Level 4 Student work and self-reflection show the student's solid skill in experimental design. Observations and data analysis indicate solid knowledge of the scientific ideas presented. Analyses and conclusions are supported by observations and data and show a moderately high level of reasoning. The entry includes solid evidence that the student worked cooperatively with others. The self-reflection sheet identifies how working with others improved the student's understanding of the scientific ideas presented. The student's investigation makes generally valid applications to real-world situations. All major aspects of the task and analysis are complete. Written expression is good and contributes to coherent and effective communication.

Level 3 Student work and self-reflection show the student's basic skill in experimental design. Observations and data analysis indicate basic knowledge of the scientific ideas presented. Analyses and conclusions may be supported by observations and data. The entry attempts to provide evidence that the student worked cooperatively with others. The self-reflection sheet identifies how working with others caused some improved understanding of the scientific ideas presented. The student attempts to make some applications to real-world situations. Some aspects of the task and analysis are incomplete. Written expression is flawed and shows basic communication.

Level 2 Student work and self-reflection show the student's limited skill in experimental design. Observations and data analysis indicate limited knowledge of the scientific ideas presented. Analyses and conclusions may not be supported by observations and data or may be incomplete. The entry may attempt to provide evidence that the student worked cooperatively with others. The self-reflection sheet may identify how working with others caused limited improvement in the student's understanding of the scientific ideas presented. When made, applications to real-world situations are limited. Tasks and analysis are largely incomplete. Written expression is flawed and shows limited communication.

Level 1 Student work and self-reflection show the student's minimal skill in experimental design. Observations and data analysis, if present, indicate minimal knowledge of the scientific ideas presented. Analyses may be incomplete and conclusions may not be supported by observations and data. The entry shows little evidence that the student worked cooperatively with others. The self-reflection sheet may identify only minimal improvement in the student's understanding of the scientific ideas presented. When made, applications to real-world situations are minimal. Tasks and analysis are mostly incomplete. Written expression is very flawed and is contributes to ineffective communication.

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by
the California Department of Education. Materials may be
copied for training purposes only.

ENTRY SCORING GUIDE

Creative Expression

Level 6 Student work and self-reflection shows high quality and attention to detail. Written explanation of work and captions show extensive knowledge of the scientific ideas presented. The student-selected format creatively presents a scientific idea in an original and unique manner that enhances the expression of the scientific idea. All aspects of the work and explanation of concepts are completed thoroughly. Written expression explaining the entry is well-organized and contributes to clear, coherent, and exemplary communication.

Level 5 Student work and self-reflection shows very good quality and attention to detail. Written explanation of work and captions show strong knowledge of the scientific ideas presented. The student-selected format presents a scientific idea in an original or unique manner that enhances the expression of the scientific idea. All aspects of the work and explanation of concepts are complete. Written expression explaining the entry is very good and contributes to clear and coherent communication.

Level 4 Student work and self-reflection shows quality and attention to detail. Written explanation of work and captions show solid knowledge of the scientific ideas presented. The student-selected format presents a scientific idea in an original or unique manner that may enhance the expression of the scientific idea. All major aspects of the work and explanation of concepts are complete. Written expression explaining the entry is good and contributes to coherent and effective communication.

Level 3 Student work and self-reflection may not show quality and attention to detail. Written explanation of work and captions show basic knowledge of the scientific ideas presented. The student-selected format may present a scientific idea in an ordinary manner that may not enhance the expression of scientific ideas. Most aspects of the work and explanation of concepts are complete. Written expression explaining the entry is flawed and exhibits basic communication.

Level 2 Student work and self-reflection shows little attention to quality and detail. Written explanation of work and captions show limited knowledge of the scientific ideas presented. The student-selected format may present the scientific idea carelessly and may not enhance the expression of the scientific idea. Aspects of the work and explanation of the concepts are largely incomplete. Written expression explaining the entry is flawed and exhibits limited communication.

Level 1 Student work and self-reflection does not display quality and attention to detail. Written explanation of work and captions show minimal knowledge of scientific ideas. The student-selected format may not present a scientific idea or the format may not enhance the expression of the scientific idea. Aspects the work and explanation of concepts are largely incomplete. Written expression explaining the entry is very flawed and exhibits ineffective communication.

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

ENTRY SCORING GUIDE

Growth Through Writing

Level 6 Student work and self-reflection show extensive knowledge of the scientific ideas presented. The student uses relevant terms appropriately, providing evidence of excellent growth in understanding of scientific ideas through comparison of original and revised samples of work. There is exceptional evidence that the entry has been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. Insightful inferences are made through the use of analogies, metaphors, examples, and personal experience. All aspects of the tasks are completed thoroughly. Written expression is well-organized and contributes to clear, coherent, and exemplary communication.

Level 5 Student work and self-reflection show strong knowledge of the scientific ideas presented. The student uses relevant terms appropriately, providing evidence of very good growth in understanding of scientific ideas through comparison of original and revised samples of work. There is convincing evidence that the entry has been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. Thoughtful inferences are made through the use of analogies, metaphors, examples, and personal experience. All aspects of the tasks are complete. Written expression is very good and contributes to clear and coherent communication.

Level 4 Student work and self-reflection show solid knowledge of the scientific ideas presented. The student uses relevant terms appropriately, providing evidence of solid growth in understanding of scientific ideas through comparison of original and revised samples of work. There is ample evidence that the entry has been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. Relevant inferences are made through the use of analogies, metaphors, examples, and personal experience. All major aspects of the tasks are complete. Written expression is good and contributes to coherent and effective communication.

Level 3 Student work and self-reflection show basic knowledge of the scientific ideas presented. The student may use terms appropriately, providing some evidence of growth in understanding of scientific ideas through comparison of original and revised samples of work. There is some evidence that the entry has been revisited and improved to reflect newly-gained knowledge, skills, and increased depth of understanding. Some inferences are made through the use of analogies, metaphors, examples, and personal experience. Most aspects of the tasks and analysis are complete. Written expression is flawed and exhibits basic communication.

Level 2 Student work and self-reflection show limited knowledge of the scientific ideas presented. The student may use terms inappropriately, providing little evidence of growth in understanding of scientific ideas. Student work may not include original and revised samples of work. There is limited evidence that the entry has been revisited and it may not have been improved to reflect any newly-gained knowledge, skills, and increased depth of understanding. Limited inferences are made through the use of analogies, metaphors, examples, and personal experience. Tasks and analysis are largely incomplete. Written expression is flawed and exhibits limited communication.

Level 1 Student work and self-reflection show minimal knowledge of the scientific ideas presented. The student may use terms incorrectly, providing little evidence of growth in understanding of scientific ideas. Student work may not include original and revised samples of work. There may not be evidence that the entry has been revisited or improved to reflect any newly-gained knowledge, skills, or increased depth of understanding. Inferences may not be made through the use of analogies, metaphors, examples, and personal experience. Tasks and analysis are mostly incomplete. Written expression is very flawed and exhibits ineffective communication.

© 1994 Golden State Examination Science Portfolio
California Department of Education

Permission for use in training has been granted by
the California Department of Education. Materials may be
copied for training purposes only

Sample 5.4

Grade 3 Interdisciplinary Task

Maryland Department of Education

Maryland is devising a set of performance tasks designed to be interdisciplinary—they require skills across content areas. The attached series of grade 3 activities (summarized from a sample task for the Maryland School Performance Assessment Program) are thematically related to the weather and require skills in science, reading, writing, and language arts. The 18 activities are given over a period of five days. Students do some work in groups and some work individually.

Steve Ferrara
Maryland School Performance Assessment Program, 1992
Maryland Department of Education
200 W. Baltimore St.
Baltimore, MD 21201
(410) 333-2369

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

Maryland School Performance Assessment Program 1992

Grade 3 Science/Reading/Writing/Language Usage

Introduction

(Students follow introduction as teacher reads aloud.)

Scientists who study the weather are called meteorologists. These scientists have been observing and recording weather information for many years. They build and use many kinds of equipment to help them measure and record their observations. By studying this information and comparing it with weather observations collected at other times of the year scientists learn about the seasons and help us understand the weather.

During the next five days you will be a weather scientist. You will be observing the weather, recording your observations on a chart you design, and studying the weather information you have collected. You will also be using a thermometer and an instrument you will be building to help you make your weather observations.

Activity 1

Your teacher will be taking you outside to observe today's weather. If you were a meteorologist or a weather scientist, how would you describe today's weather to another scientist? When you return to your classroom list at least four features of today's weather.

Activity 2

Soon you will be starting to make daily observations of the weather. You need to design a chart to use for recording your observations. Now work with a group of your classmates to list below all the different kinds of weather information that you might want to include on your chart. You will have a thermometer to use to include temperature on your list.

Activity 3

Now, working alone, design a chart or table that you will use to record your weather observations for the next five days. Include on your chart at least four weather features from your group list.

Bridging Activity

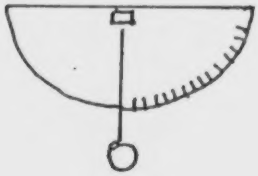
You might have included wind speed on your chart. If not, add wind speed to your chart now. Then look at these instructions for building an anemometer and answer the questions that follow.

Say: When you use directions to build something, it is a good idea to read and think about them carefully before you start. Here are some directions for building an anemometer and some questions to answer before using the directions.

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

To make your anemometer do these things:

1. Tie a knot in one end of your string.
2. Measure 30 centimeters from the knot and tie another knot.
3. Place the other knot against the foam ball and tape the string to the ball.
4. Let the ball swing freely in the wind and observe the line on the anemometer outline where the string crosses. This is the number you can use to find the wind speed.



30 cm from knot to knot

#	Wind Speed
1	6
2	8
3	10
4	12
5	14
6	16
7	18
8	20

Activity 4 Global Understanding

Why are anemometers useful for observing the weather?

Activity 5 —Developing Interpretation

What is the purpose of the foam ball?

Activity 6

What is the relationship between the wind and the position of the string?

Activity 7—Developing Interpretation/Personal Response/Critical Stance

Are the directions clear enough? Explain why or why not.

Activity 8— Developing Interpretation/Critical Stance

Why do you think the author showed a chart with the picture of the anemometer?

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

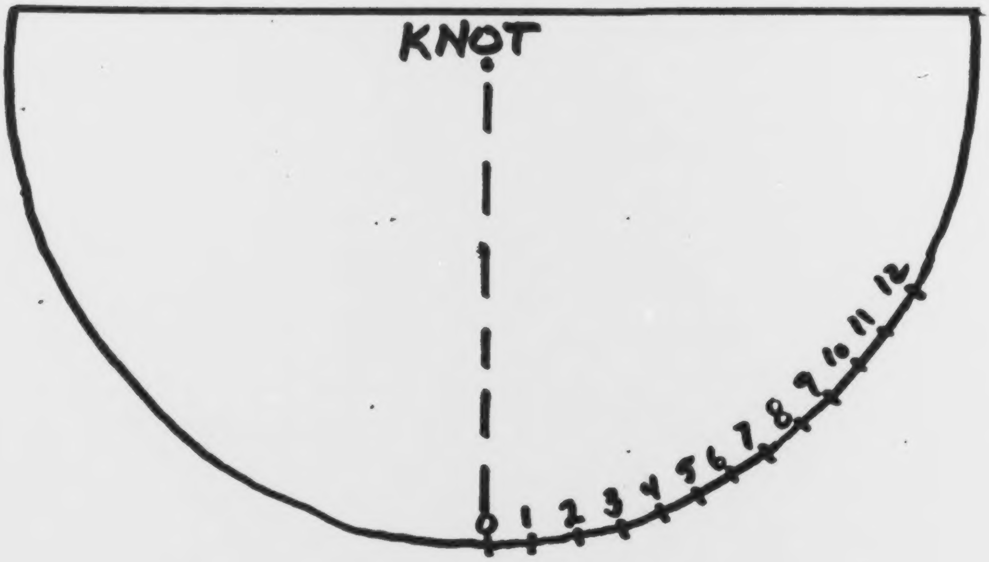
Activity 9—Developing Interpretation/Personal Response

Is there any other information you think you would need to build and use this anemometer?

Activity 10

Now, here is a different set of directions for making an anemometer. Read these new directions. Then, use either set of directions or both sets of directions to build a wind vane.

1. Cut out the anemometer outline and glue it to a piece of cardboard.
2. Tie a knot in one end of a piece of string.
3. Measure 30 centimeters from the knot and tie another knot.
4. Place one knot against the polystyrene ball and tape the string to the ball.
5. Place the other knot on spot marked "KNOT" on the anemometer outline and tape it to the outline. The string should follow the line between the "KNOT" and the "0" on the outline.
6. Aim you anemometer into the wind so that the ball swings freely. Read the number that the string crosses. Then use the chart to get the wind speed in miles per hour.



Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

Activity 11

Your teacher wants to know how you built your anemometer. Write a report for your teacher identifying which set of directions you used or whether you used both set of directions. Explain the reasons for the choice you made. Tell in your report what was helpful in each set of directions and what was a problem in each set of directions. Because your report will be read by your teacher, be sure it is clear and complete. Also, check for correct spelling, punctuation, grammar, and capitalization.

Activity 12—Language Usage

Your friend is confused about how to build an anemometer. In order to help your friend you have decided to write your own set of directions for building an anemometer which will be easy for your friend to follow. Use information from either or both sets of directions that were given to you and what you learned as you built your own anemometer in order to help you write a set of clear directions for your friend. Because the directions you write will be read by your friend, be sure your directions are clear and complete. Also, check for correct spelling, punctuation, grammar, and capitalization.

Activity 13—Personal Response/Critical Stance

How is the model you built different from the illustrations of the anemometer in the sets of directions?

Activity 14—Personal Response/Critical Stance

Each day for the next five days go outside to make your individual weather observations including temperature and wind speed and then record them on your chart. Finish all of your observations before continuing this task.

Activity 15—Personal Response/Critical Stance

Last year the students decided that it was important to make their observations at the same time each day. Explain why this might be important. Include an example of what might happen if each day you changed the time of your observations.

Activity 16

Select one of the weather features you have been observing and make a graph of the information you have collected.

Activity 17

Get together with members of your group. Compare your weather observations and measurements with those of other members of your group. Are any of your results very different from someone else's? If so, describe that difference and explain why that might have happened.

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

Activity 18

The local television station has invited you to represent your class and appear on television to inform viewers about what you have learned about the weather. Write a script that you will read informing the television audience about the weather conditions you have observed and recorded. Be sure to explain how the weather has been the same or different from what you would usually expect for this time of year. Also, use information from the anemometer you made and from the observations you recorded on your weather chart as you write an interesting and informative script to read for your television audience.

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

SAMPLE SCORING TOOLS

Grade 3: Science/Reading/Writing/Language Usage

Activity 1
Science Outcome 5: Processes of Science -observing

2= The response thoroughly demonstrates the ability to observe and identify weather features. It does this by listing at least four specific features, with some degree of quantification.

1= The response adequately demonstrates the ability to observe and identify weather features. It does this by listing four general or three specific features, with little or no quantification.

0= Other

Activity 3
Science Outcome 5: Processes of Science-communicating

2= The response thoroughly demonstrates the ability to communicate findings in graphic form. It does this by presenting a chart with four or more features on one axis and five days on the other; both axes are clearly labeled.

1= The response adequately demonstrates the ability to communicate findings in graphic form. It does this by presenting a chart with two-three features on one axis and both axes correctly labeled or four features, with one axis not clearly or correctly labeled.

0= Other

Activity 4
Reading to Perform a Task: Global Understanding

2= Syntheses of multiple ideas or one idea extended.

1= One overly general or overly specific idea.

0= Other

Activity 7
Reading to Perform a Task: Developing Interpretation/Personal Response/Critical Stance

3= Opinion stated and supported with multiple ideas from the text and substantial inferences supported by references to the text.

2= Opinion stated and supported with some ideas from the text and some inferences supported by references to the text.

1= Opinion stated with general references to the text or opinion implied with some support.

0= Other

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

Activity 11
Reading to Perform a Task: Developing Interpretation/Personal Response/Critical Stance

3= Opinion stated and supported with multiple ideas from the text and substantial inferences supported by references to the text; more than one perspective evident.

2= Opinion stated and supported with some ideas from the text and some inferences supported by references to the text; more than one perspective evident.

1= Opinion stated with general references to the text or opinion implied with some support; only one perspective evident.

0= Other

Activity 12
Language Usage

2= Consistently uses word and sentence order and language choices to express meaning with style and tone. Text conveys uniform impression of correctness* and any errors that are present represent risk-taking.

1= Sometimes uses word and sentence order and language choices to express meaning with style and tone. Text conveys uniform impression of correctness and any errors that are present represent risk-taking.

0= Rarely or never uses word and sentence order and language choices to express meaning with style and tone. Text appears error-ridden.

* correct usage, punctuation, spelling and capitalization.

Activity 16
Science Outcome 5/Processes of Science-communicating

2= The response thoroughly demonstrates the ability to communicate findings in graphic form. It does this by presenting a chart that represents collected data accurately and completely.

1= The response adequately demonstrates the ability to communicate findings in graphic form. It does this by presenting a chart that represents collected data with minor inaccuracies and/or omissions.

0= Other

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

PROMPT

SCORING RUBRIC: WRITING TO INFORM

3 points: Development: The writer provides accurate, specific purposeful information that is extended and expanded to fully explain the topic.

Organization: The writer establishes an organizational plan and consistently maintains it.

Attention to Audience: The writer provides information relevant to the needs of the audience.

Language: The writer consistently uses language choices to enhance the text.

2 points: Development: The writer provides information that adequately explains the topic with some extension of ideas. The information is usually accurate and purposeful.

Organization: The writer establishes and maintains an organizational plan, but the plan may have some minor flaws.

Attention to Audience: The writer provides information relevant to the needs of the audience.

Language: The writer consistently uses language choices to enhance the text.

1 point: Development: The writer provides information that inadequately explains the topic, with some extension of ideas. The information is sometimes inaccurate, general, or extraneous.

Organization: The writer generally establishes and maintains an organizational plan.

Attention to Audience: The writer provides some information relevant to the needs of the audience.

Language: The writer sometimes uses language choices to enhance the text.

0 points: Development: The writer provides insufficient information to explain the topic. The information provided may be vague or inaccurate.

Organization: The writer either did not establish an organizational plan or, if an organizational plan is established, it is only minimally maintained.

Attention to Audience: The writer did not provide information relevant to the needs of the audience.

Language: The writer seldom, if ever, uses language choices to enhance the text.

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

Codes: **A:** Blank - There is no response.

B: The writer's response is off-task and/or off-topic. It does not address the question that was asked.

C: Unscorable - The writer's response cannot be read (e.g., it is illegible, incomprehensible).

D: Copied from test text.

LANGUAGE IN USE RUBRIC

3 Points: Throughout the entire text there is evidence of the writer's using language choices and order for effective style, tone, and expression of meaning. The text uniformly conveys an impression of correctness, with few, if any, errors. Errors that occur may appear as a consequence of risk-taking in language use. The writer consistently:

- uses varied sentence formation to create style and tone and to enhance meaning
- uses language choices to create style and tone and to enhance meaning
- demonstrates correct usage, punctuation, spelling, and capitalization

2 Points: Through much of the text there is evidence of the writer's using language choices and order for effective style, tone, and expression of meaning. The text generally conveys an impression of correctness. The errors that occur may be of one or two types and occur infrequently. Sometimes errors that occur may appear as a consequence of risk-taking in language use. The writer frequently:

- uses varied sentence formation to create style and tone and to enhance meaning
- uses language choices to create style and tone and to enhance meaning
- demonstrates correct usage, punctuation, spelling, and capitalization

1 Point: In portions of the text there is evidence of the writer's using language choices and order for effective style, tone, and expression of meaning. Errors of several types may occur, and may be repeated. Errors do not appear to be the result of risk-taking. The writer sometimes:

- uses varied sentence formation to create style and tome and to enhance meaning
- uses language choices to create style and tone and to enhance meaning
- demonstrates correct usage, punctuation, spelling, and capitalization

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

0 Points: In little or none of the text is there evidence of the writer's using language choices and order for effective style, tone, and expression of meaning. The text conveys an overall impression of being error-ridden. Errors of several types may occur repeatedly. Errors do not appear to be the result of risk taking. The text may be so brief that there is limited evidence of correct language use. The writer rarely or never:

- uses varied sentence formation to create style and tone and to enhance meaning
- uses language choices to create style and tone and to enhance meaning
- demonstrates correct usage, punctuation, spelling, and capitalization

Codes: **A:** Blank - There is no response.

B: The writer's response is off-task and/or off-topic. It does not address the question that was asked.

C: Unscorable - The writer's response cannot be read (e.g., it is illegible, incomprehensible).

D: Copied from test text.

Activity 18

Science Outcome 1: Earth Science

3= The response thoroughly demonstrates the acquisition and integration of major concepts and unifying themes from earth/space sciences. The response does this by thoroughly demonstrating the understanding that monitoring the weather allows us to learn more about the characteristics of the seasons.

2= The response generally demonstrates the acquisition and integration of major concepts and unifying themes from earth/space sciences. The response does this by generally demonstrating the understanding that monitoring the weather allows us to learn more about the characteristics of the seasons.

1= The response adequately demonstrates the acquisition and integration of major concepts and unifying themes from earth/space sciences. The response does this by adequately demonstrating the understanding that monitoring the weather allows us to learn more about the characteristics of the seasons.

0= Other

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

The 1992 Maryland School Performance Assessment Program:

Scoring Student Responses

As in 1991, the 1992 Maryland School Performance Assessment Program (MSPAP) is made up of performance tasks. Students' response to these tasks are "open-ended"—that is, students generate their own responses. A wide range of responses to each task is acceptable. Some responses receive full credit, depending upon the quality of the response. Responses to MSPAP tasks are scored by trained readers, unlike multiple choice items which can be scored by machine. One of three kinds of scoring tools is used to assign scores to student responses in the MSPAP:

Scoring Rubric: A score scale (the number of points that can be given) and set of descriptions of response characteristics and quality for each score point. MSPAP rubrics describe what students know and can do in terms of particular outcomes for each of the MSPAP content areas. A rubric may be used to score responses or may serve as a "blueprint" for briefer scoring tools that fit particular activities. The Maryland Writing Test and other essay tests are scored using rubrics.

Scoring Rule: A smaller score scale and brief set of descriptions for each score point which might be considered a condensed rubric.

Scoring Key: An activity-specific score scale and set of descriptions.

Scoring rubrics and rules may be used to score responses to different activities in a MSPAP content area. Scoring keys apply to single assessment activities. Regardless of which scoring tool is used, the number of levels of performance likely for that activity determines the number of possible score points in the tools. MSPAP scoring tools may have as many as seven score points (0, 1, 2, 3, 4, 5, 6) or as few as two (0, 1).

Scoring rubrics are used for: a) writing; b) extended reading responses; c) responses intended to measure the mathematics strands of problem-solving, communication, reasoning, and connections; and d) some science and social studies responses.

Typically, a scoring rule is used to score brief responses for language usage, a single reading stance or mathematics content area, and some science or social studies activities.

Scoring keys are used most often for scoring responses when specific product or range of information is sought. Unlike some "answer keys," however, MSPAP scoring keys reflect the language of the State Board-adopted outcomes for the content area being assessed.

Other resources that are used to help make score decisions include sample responses for which "true scores" have been reached by consensus by Maryland educators. These are call "rangefinding" or "anchor" responses. They are models of responses at each score point of scoring tool.

Integration of content areas in MSPAP tasks requires that some responses be scored multiple times for different content areas and outcomes. For example, essays in the MSPAP are scored once for a writing purpose and once for language usage. Similarly, other responses may be scored once to contribute to a mathematics outcome score and a second time to contribute to a science outcome score. This process is referred to as "successive" scoring. Sometimes a single key or rule may be applied which permits the measurement of outcomes from more than one content area. This process is referred to as "simultaneous" scoring. The sample integrated tasks provide examples of both successive and simultaneous scoring for different content areas. In successive scoring the score in one area does not impact the score in the other. Also, successive scores are not given by the same individual, except in the case of writing and language usage scores for responses to writing prompts.

Permission for use in training has been granted by the Maryland Department of Education. Materials may be copied for training purposes only.

Sample 5.5

Assessment of Learning and Communication Processes

Alberta Education

The goals for the *Evaluating Students' Learning and Communication Processes* program are to : (1) evaluate progress of secondary students (grades 7-10) in six learning and communication processes; (2) integrate the six processes across classes in language arts, social studies, and science; and (3) empower students to take control of learning by making them conscious of the six process skills and how they, themselves, use them.

There is a general handbook for all subject areas that covers evaluation, performance criteria, recording information, and instruction (how to implement the program, instructional activities for students, help with student self-reflection, help with teacher collaboration, and how to report student progress). There is a separate handbook for each subject area that contains sample teaching units designed to show teachers how to incorporate diagnostic evaluation of students' learning and communication processes into regular instruction.

The attached pages (from Handbook 1) contain information about the sample instructional units and performance criteria. They are reproduced with the permission of Alberta Education.

From *"Evaluating Students' Learning and Communication Processes," 1994*
Alberta Education
Learning Resources Distributing Centre
12360 142nd St.
Edmonton, AB T5L 4X9, Canada
(403) 427-2767

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

The Diagnostic Teaching Units

Five diagnostic teaching units have been designed to help teachers implement *Evaluating Students' Learning and Communication Processes* in their classrooms. These are found in three separate handbooks called

Diagnostic Teaching Unit: Language Arts

Diagnostic Teaching Units: Social Studies

Diagnostic Teaching Units: Science

Designed by teachers to show you how to include diagnostic evaluation of learning and communication process in your program

These units have been created by teachers and tested in their classrooms. They are based on the programs of study for Alberta secondary schools. The unit and lesson plans for each diagnostic teaching unit show you how to incorporate diagnostic evaluation of students' learning and communication processes as part of your regular instructional program.

Examples showing you how to analyse and evaluate learning and communication processes in student work

The unit and lesson plans are followed by an analysis of student work produced when the units were tried out in classrooms. This will help you learn how to identify and evaluate students' use of the six learning and communication processes in different kinds of student work. It will also help you select appropriate instructional activities to help students develop more independent control in their use of these processes where this is required. The five diagnostic teaching units are

Grade 7 Language Arts: What's So Funny? (The Lighter Side of Life)

Grade 7 Social Studies: Cultural Transition in Japan

Grade 7 Science: Structures and Design

Grade 8 Social Studies: The Physical and Human Geography of the Amazon Basin

Grade 9 Social Studies: Soviet Leadership and Economic Growth

Grade 10 Science: Acids and Bases

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

A model for designing your own diagnostic teaching units

In addition to helping you implement the program in your classroom, these units also provide a model to assist you in planning your own diagnostic teaching units. In Appendix 2 of this handbook, you will find unit planning and lesson planning sheets, which you may photocopy to use when planning your own diagnostic teaching units. In a later section of this handbook, you will find further information to help you design your own diagnostic teaching units.

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

DESCRIPTIVE SCALE
EXPLORING

	Key Indicators			
	becomes aware of prior knowledge, feelings, and values	**frames questions; searches for additional information**	**connects new with prior knowledge, feelings, and values**	**takes calculated risks where appropriate**
Consistently Independent	offers variety of ideas that reflect personal knowledge, feelings, and values	frames appropriate questions and searches a variety of sources	considers own position in light of new ideas	acts on inconsistent or incomplete data where appropriate
Frequently Independent	offers some ideas that reflect personal knowledge, feelings, and values	frames questions and searches available sources but may require reassurance	considers some aspects of own position in light of new ideas	acts on inconsistent or incomplete data where appropriate but may require reassurance
Needs Some Assistance	with encouragement, offers some ideas that reflect personal knowledge, feelings, and values	with help, frames questions and searches readily available sources	with help, considers some aspects of own position in light of new ideas	with help, acts on inconsistent or incomplete data where appropriate
Needs Much Assistance	even with encouragement, has difficulty offering ideas that reflect personal knowledge, feelings, and values	even with help, frames questions and searches haphazardly, using only what is at hand	even with help, considers new ideas only when they fit personal beliefs and opinions	even with help, has difficulty acting on inconsistent or incomplete data where appropriate

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

DESCRIPTIVE SCALE
NARRATING

	Key Indicators			
	uses time and space to organize remembered experience and information	**relates experience within and across subjects to clarify concepts**	**uses anecdotes in sharing experience**	**values and enjoys sharing experience, real and vicarious**
Consistently Independent	shares experience and information in a consistently coherent way	transfers relevant information and experience across topics	recalls, creates, or adapts anecdotes that clearly support meaning	indicates enthusiasm
Frequently Independent	shares experience and information in a generally coherent way	transfers some information and experience across topics	recalls, creates, or adapts anecdotes that generally support meaning	indicates interest
Needs Some Assistance	with some help, shares experience and information	with help, transfers some information and experience across topics	with prompting, shares anecdotes that generally support meaning	with prompting, indicates acceptance
Needs Much Assistance	even with help, shares experience and information randomly	even with help, has difficulty understanding relevance of information from other topics	even with prompting, seldom shares anecdotes that support meaning	even with help, indicates impatience or frustration

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

DESCRIPTIVE SCALE
IMAGINING

	Key Indicators			
	creates images and conveys associated feelings	**transforms images**	**imagines self in different situations, places, or times**	**uses figurative language**
Consistently Independent	selects effective detail to appeal to more than one sense; conveys feelings associated with images	alters and elaborates images to enhance meaning	imagines self in a variety of unfamiliar contexts	uses figurative language to enhance meaning
Frequently Independent	selects some effective detail to appeal to one or more senses; conveys some feelings associated with images	alters or elaborates images to clarify meaning	imagines self in some unfamiliar contexts	uses some figurative language to enhance meaning
Needs Some Assistance	with help, selects some detail to appeal to only one sense, usually vision; with prompting, conveys feelings associated with images	with help, alters or elaborates images to clarify meaning	with help, may imagine self in some unfamiliar contexts	with help, uses some figurative language to enhance meaning
Needs Much Assistance	even with help, has difficulty selecting detail or conveying feelings associated with images	even with help, has difficulty altering or elaborating images to clarify meaning	even with help, has difficulty imagining self in unfamiliar contexts	even with help, has difficulty using figurative language to enhance meaning

Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 5.5: *Assessment of Learning and Communication Processes*
LNP Contributor: Judy Arter, NWREL

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

DESCRIPTIVE SCALE
EMPATHIZING

	Key Indicators			
	shifts attention away from self while communicating	**reserves judgment and disbelief where appropriate**	**selects language that takes audience into account**	**takes on the role of another**
Consistently Independent	actively attends to others and encourages communication	reserves judgment and disbelief in most situations where appropriate	varies language to suit many different situations	uses language and action that consistently represent the appropriate role
Frequently Independent	attends to others and receives communication openly	reserves judgment and disbelief in some situations where appropriate	varies language to suit familiar situations	uses language and action that suggest the appropriate role
Needs Some Assistance	with help, attends to others and receives some communication openly	with help, reserves judgment and disbelief in some situations where appropriate	with help, varies language to suit familiar situations	with help, uses language or action that suggest the appropriate role
Needs Much Assistance	even with help, has difficulty attending to others; receives communication compatible only with personal interests	even with help, has difficulty reserving judgment and disbelief where appropriate	even with help, has difficulty varying language according to situation	even with help, has difficulty using language or action that suggest the appropriate role

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

DESCRIPTIVE SCALE
ABSTRACTING

	Key Indicators			
	supports generalizations	**applies generalizations**	**evaluates the soundness and significance of generalizations**	**uses symbols**
Consistently Independent	provides relevant support for generalizations	makes plausible predictions and gives convincing explanations	examines generalizations from several perspectives	provides effective representation of a complex idea
Frequently Independent	provides some support for generalizations	makes some predictions and gives some explanations	examines generalizations from an alternate perspective	provides representation of a complex idea
Needs Some Assistance	with help, provides some support for generalizations	with help, makes predictions and gives explanations	with help, examines generalizations from an alternate perspective	with help, provides representation of a complex idea
Needs Much Assistance	even with help, has difficulty providing support for generalizations	even with help, has difficulty making predictions and giving explanations	even with help, has difficulty examining generalizations from an alternate perspective	even with help, has difficulty providing representation of a complex idea

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

DESCRIPTIVE SCALE
MONITORING

	Key Indicators			
	sets goals for learning and communication	**plans strategies for learning and communication**	**adjusts goals and strategies for learning and communication**	**facilitates learning and communication and perseveres**
Consistently Independent	sets realistic goals relevant to purpose	selects strategy relevant to purpose from a variety of possibilities	adjusts goals and strategies to fit changing communication context	elicits and maintains interest and momentum in spite of difficulties
Frequently Independent	sets some realistic goals related to purpose	selects strategy related to purpose from a set of possibilities	adjusts goals and strategies to take into account some aspects of changing communication context	maintains interest in spite of difficulties
Needs Some Assistance	with help, sets some realistic goals related to purpose	with help, selects strategy related to purpose from a set of possibilities	with help, adjusts goals and strategies to take into account some aspects of changing communication context	with prompting, maintains interest
Needs Much Assistance	even with help, has difficulty setting realistic goals related to purpose	even with help, has difficulty selecting strategy related to purpose	even with help, has difficulty adjusting goals and strategies to take into account changing communication context	even with prompting, has difficulty maintaining interest

Permission for use in training has been granted by Alberta Education. Materials may be copied for training purposes only.

Sample 5.6

ALGEBRA II
MICHAEL LEHMAN, HOLT PUBLIC SCHOOLS

The author has attempted to devise a better method of assessing algebra problem solving, concepts, and skills than traditional paper and pencil tests. Students work together in small groups to solve six problems. They then explain their solutions in front of a panel of judges which can request any member of the group to explain any problem. The group then is assigned new problems to solve in front of the judges.

Attached are instructions for raters, the scoring procedure with some assistance on how to assign grades, instructions for the students, and one of the six Algebra II problems to be solved in the group.

Michael Lehman
"Algebra II," 1993
Holt Public Schools
1784 Aurelius Road
Holt, Michigan 48842
(517) 694-2162

Permission for use in training has been granted by Michael Lehman. Materials may be copied for training purposes only.

Algebra II
Performance Assessments

Thank you for your willingness to help with our performance assessment. I hope the following information will be helpful to you as you plan to be a judge for the upcoming performance assessment. This performance assessment will take place on Wednesday, June 3rd during the 4th hour exam period (9:15 to 11:50). Please be prompt so that the students will have ample time to discuss their problems with you. I will try to get the students organized as much as I can the day prior to the assessment. It will take a few minutes to take attendance and send them to you.

Please keep in mind that for a lot of the students this is still a very high stress experience. They have experienced this three times before in the form of presenting portfolio problems they prepared in the fall, their semester exam in January, and their quarterly exam in March, so please do push them a little so they get a clear picture of what they do understand and what they don't. Please give the students plenty of opportunity to explain themselves, but if it is obvious that they are trying to fake it or are unsure of themselves, let them know that it is not what we are after and move on. Do not allow them to read prepared scripts. I am interested in their thinking and in their ability to explain, not in their ability to read. The students will have had an opportunity to work on the problems in class and to practice explaining them, so this should not be a problem.

Only one student per problem. They have been instructed that they will have to discuss the problem on their own without help form other members of the group. *They have also been told that the judges will pick the problem for them instead of them choosing.* After you feel this student is finished you may ask another student some questions about the same problem as a follow-up or to verify the previous student's explanation. Time has always been a problem and probably will continue to be one. Please keep track of the time so you can give all the students a fair chance and not have to rush the last student

I am trying something a little different this time. The students will have prepared six problems for this exam. You should use the first hour to allow each student to present one problem. You will then be given four additional problems of which you can pick one or more and have the students as a group explain to you how they would solve it. They do not have to do the actual computations unless you want them to because this will slow down the process. All they need to do is explain how they would solve it to your satisfaction that if necessary they could find the final results. I hope this will give additional information about the student's understanding and their ability to apply the mathematics we have discussed.

Pick one problem from the packet of problems. Ask the student to explain the problem. You may want to use some of the following questions to help focus the discussion.

What is the problem about?

What is the problem asking for?

What strategy did you use to solve the problem?

What in the problem made you think to use your strategy?

What were your results?

Where did the formula you used come from? (If a particular formula is used)

Why does this formula work? What made you think to use it?

How do you know your results are correct? ("Because they are the same as the way the book did it" or "That is how we did it in class" is not enough.)

Permission for use in training has been granted by Michael Lehman. Materials may be copied for training purposes only.

You are also free to ask any other questions you feel are appropriate for the discussion. Please pursue points and details with the students and search for their understanding of the problems. If a student uses terminology that you feel needs explanation please ask them for an explanation. For example such terms as function, directrix, parabola, logarithm, etc. Make sure they know what they are talking about and can explain it.

If a student seems unprepared, let them do what they can and then move on. If time permits you may want to come back and let students clarify any points they may want to after having some time to think about the problem further. Please make note of this on the evaluation form. After all students have had a chance to discuss a problem, go on to the new problems.

Please use the following evaluation sheet in assessing the students' discussions. If you find the categories I have outlined unusable or too constraining please write comments in the comment section or on the back. Once again I am including the suggestions for grading criteria. In assigning the final points you need to be as specific in your comments as possible. Remember that I will need your comments to discuss the students' evaluations with them. The students find your comments very interesting and are anxious to read them. In order for them to be useful to the students, please be as detailed as possible. Short phrases or copies of student computation do not give the students enough information in order for them to improve.

You will find copies of the problems, the evaluation sheet, and the suggested grading-criteria in this packet. I will provide copies of the evaluation sheet for each student during the assessment. Please remember that these are only suggested answers. If a student interpreted the problem differently and can defend their interpretation then they can have different solutions. You will need to decide if their interpretation and rationale is appropriate.

Thank you for all your help. You have truly helped my students learn and enhance their understanding of what it means to know and understand mathematics. Without you none of this would be possible.

Judges Teams:

Team #1
Perry Lanier
Trudy Sykes
Chet Franck

Documentor
Jessica Zimmerman

Team #2 (Library)
David Mucznski
Samuel LoPresto
Mark Maksimowicz

Documentor
Randy LaFeve

Team #3 (Library)
Sandy Bethell
Ron Van Ermen
Tom Bird

Documentor
Katie Nott

Team #4(Library)
Dan Chazan
Scott Szpara
Chan Nauts

Documentor
Jeff Milbourn

Team #5(Room 214)
Bill York
Kathy Burgess
Debbie Roeske

Documentor
Steve Streeter

Team #6(Room 214)
Steve Kersner
Patti Summers
Linda Alford

Documentor
Beth Bonner

Team #7(Principal's Conference Room)
Tom Davis
Ted Gardella
Jackie Wood

Documentor:
Laura Hendricks

Permission for use in training has been granted by Michael Lehman. Materials may be copied for training purposes only.

Algebra II
Discussion Final

Name:______________________

Presentation of Prepared Problems

Mathematics:

1) Making Sense of Problem (Understanding Concepts) 1 2 3 4 5

2) Problem Solving Strategies (Methods Used) 1 2 3 4 5

3) Accuracy of Results 1 2 3 4 5

4) Interpreting Results (What Do the Results Mean?) 1 2 3 4 5

Clarity of Explanation:

1) Ability to Communication Results (Clarity, Use of Charts/Graphs) 1 2 3 4 5

2) Explanation (Able to Answer Questions) 1 2 3 4 5

Discussion of Group Problems

1) Contributed ideas towards the solution of the problem. 1 2 3 4 5

2) Group was able to solve the problems presented with this student's help. 1 2 3 4 5

3) With this student's help the group was able to explain their method of solution to the judges in a way that helped the judges to understand the mathematics involved. 1 2 3 4 5

4) This student demonstrated to the judges that he/she understands the mathematics involved in this situation. 1 2 3 4 5

Overall Score __________

Laboratory Network Program—*Alternative Assessment Toolkit*
Sample 5.6: *Algebra II*
LNP Contributor: Judy Arter, NWREL

Permission for use in training has been granted by Michael Lehman. Materials may be copied for training purposes only.

Algebra II
Performance Assessment
Grading Criteria

Thoughts about grading:

The following are suggestions to help you in your grading of the students. You may use them as guidelines or you may choose to set up your own guidelines.

Suggestions for an A:

Students receiving an A should be able to demonstrate to you that they have a clear understanding of the problem and all the concepts it contained. They should be able to make sense of their results in relationship to the situation given. They should be able to clearly communicate their understanding to you.

Suggestions for a B:

Students receiving a B should be able to demonstrate a good understanding of the problem and the concepts it contained. They should be able to make sense of their results in relationship to the situation given. They should be able to communicate their understanding to you although it may not be as clear as you would like. The difference between an A and a B would be in the confidence the students shows in their work as well as the level of understanding they demonstrate.

Suggestions for a C:

Students receiving a C should be able to demonstrate an adequate understanding of the problem and the concepts it contained. Their understanding may not be as complete as in an A or a B, but adequate enough to give you confidence that they understand what you feel are the important concepts. They may have some trouble making sense of the results but are able to do so with some probing from you.

Suggestions for a D:

Students receiving a D would demonstrate a lack of understanding of some of the key concepts contained in the problem. They would seem to be able to go through the motions to get the results but are unable to explain why they solved it the way they did, other than to say "That's how we did it in class." They are also unable to make much sense of their results even with some probing from you.

Suggestions for an E:

Students receiving an E would demonstrate a clear lack of understanding of most of the key concepts contained in the problem. They would be unable to explain their results and why they solved the problem the way they did. They would give you a feeling that they have no understanding of the mathematics involved. The student may appear to have done little preparation for this type of assessment.

Permission for use in training has been granted by Michael Lehman. Materials may be copied for training purposes only.

Algebra II
Discussion Final
June 1992

Name______________________________
Hour_________

You should find six problems included in this packet. You should work on these problems as a group as well as on your own time. During the exam period you will be asked to explain your results before a panel of judges. Each member of the group must be able to explain each problem by themselves, as the judges will pick the problem to be presented. Other members will be present but will not be able to offer ideas on an individual's problem. Since you will not know which problem you will be asked, be sure to study all the problems. In your explanations include samples of graphs you may have used, calculations you may have done, charts you made up and any other information you feel will help the judges understand what you know. **Do not write a script** that you intend to read as this would only prove that you can read.

You will have approximately one hour to present your problems. During the last 1/2 hour of the exam period the judges will give you one to three problems that you have not seen before. You will be asked to explain how you would solve these problems. You will not have to do all the computation unless you feel it will add to your explanation or the judges feel they need them to understand what you are saying.

Suggestion for Exam Day:

1) Wear nice looking clothes.

2) Look the judges in the eye and talk directly to them.

3) Say "Please" and "Thank you" when you ask for a questions to be repeated and when you are finished.

4) Don't say "I don't know."

5) If you don't understand a question, ask for clarification or that they ask the question another way.

6) If you still don't know, answer another question! By going on, you may be able to figure out the answer to the original question.

7) If you find yourself becoming frustrated take a deep breath or ask if you could have a few minutes to think about your answer.

8) Have confidence in yourself. If you have prepared properly, you should be able to handle any questions that come up. Remember, you are the authority.

Permission for use in training has been granted by Michael Lehman. Materials may be copied for training purposes only.

Sample Problem

1) You and your partner have decided to go looking for a buried treasure described on a scrap of paper found in the basement of an old house. The only clues to the treasures location is the following:

"The treasure is buried in a spot that is the same distance from the boulder as it is from the railroad tracks. It is also..."

And the rest of the information is missing. But some other clues you may be wise to consider are:

1) the distance from the tracks to the boulder is 11 yards.

2) consider the tracks as a straight line.

3) keep all of the units in yards or feet.

Keep in mind that the distance of the treasure from the railroad track is interpreted as being the length of a perpendicular drawn to the tracks from the treasure.

Explain how you can find all the possible places where the treasure may be buried. Use the information given above to demonstrate your conjecture.

Sample 5.7

Mathematics Assessment

California Department of Education

As described in detail in the *California Mathematics Framework* (1992), mathematically powerful students are those who can draw on mathematical ideas, tools, and techniques to think and communicate. In responding to an open-ended question and accomplishing its task, a student demonstrates mathematical power. As the student engages in and responds to the task, he or she draws from his or her thinking capacity, understanding of mathematical ideas, ability to use tools and techniques, communication skills, and ability to shape a coherent and focused response.

Teachers use a rubric to score and evaluate students' responses to open-ended questions or investigations. The scoring rubric, based on the goals articulated in the *Mathematics Framework*, describes how well the student work meets the expected standard of performance. Teachers should study the general rubric as well as the evaluation process to understand how mathematical power is assessed through open-ended mathematical problems.

Table 1 shows a general rubric used to evaluate responses to open-ended questions in mathematics. The rubric articulates, at six levels, the extent to which student work accomplishes the purpose of the task and demonstrates mathematical understanding, reasoning, thinking, communicating, and use of tools and techniques. Level 6 represents the highest quality of work, and level 1, the lower quality of performance.

This rubric can be applied to any open-ended task. Therefore, before applying this rubric to mathematics assessment, the scorer must explore all possible ways in which it relates to a particular problem. In other words, the scorer looks for the mathematical ideas, thinking, communication, tools and techniques that a student can use to solve a particular problem.

"A Sampler of Mathematics Assessment—Addendum"
California Department of Education
721 Capital Mall
Sacramento, CA 94244

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Table 1. General Rubric

Level 6	**Solid work that may go beyond the requirements of the task(s), showing for example;** • complete understanding of the task's mathematical concepts and processes. • clear identification of all of the important elements of the task(s). • where appropriate, clear evidence of doing purposeful mathematics, including investigating, experimenting, modeling, designing, interpreting, analyzing, or solving. • excellent prose and mathematical supporting arguments that may include examples or counter-examples. • creativity and thoughtfulness in communicating the results and the interpretations of those results, to an identified audience, using dynamic and diverse means. • multiple solutions based upon different assumptions about or interpretations of the task(s). • unusual insights into the nature of and the resolution of problems encountered in the task(s). • a high level of mathematical thinking that includes, where appropriate, making comparisons, conjectures, interpretations, predictions, or generalizations. • exceptional skill in choosing appropriate mathematical tools and techniques in the resolution of problems in task(s).
Level 5	**Fully achieves the requirements of the task(s), showing for example;** • good understanding of the task's mathematical concepts and processes. • identification of most, if not all, of the important elements of the task(s). • evidence of doing purposeful mathematics, including where appropriate, investigating, experimenting, modeling, designing, interpreting, analyzing, or solving. • clear, successful communications with an identified audience. • one solution and interpretation of those results. • evidence of mathematical thinking that includes, where appropriate, making comparisons, conjectures, interpretations, predictions, or generalizations. • use of variety of tools and techniques appropriate to the form of the task(s) and the requirements of the task.

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Table 1. General Rubric *(continued)*

Level 4	**Substantially completes the requirements of the task(s), showing for example;** • an understanding of most of the task's mathematical concepts and processes. • identification of the important elements of the task(s), but some less important ideas are missing. • some aspects of investigations, experiments, model building, designs, interpretations, analysis, solutions required by the task(s) may be missing, but most of the parts are included. • adequate communication with an identified audience, but with limited clarity and variety. • occasional evidence of mathematical thinking involving comparisons, conjectures, interpretations, predictions, or generalizations. • a limited variety of tools and techniques used to resolve the situation presented in the task(s).
Level 3	**Limited completion of the requirements of the task(s), showing for example;** • an understanding of some of the task's mathematical concepts and processes, but with evidence of gaps in those understandings. • identification of some of the important elements of the task(s), but assumptions about some of the elements may be flawed. • communication of some ideas, but generally makes inadequate attempts to communicate, often failing to address the identified audience, and difficulty in expressing mathematical ideas. • inadequate mathematical thinking that includes ineffective analysis procedures, limited solution strategies, unclear mathematical arguments, and inappropriate interpretation of results. • a selection of some inappropriate tools and techniques used to resolve the situation presented in the task(s).

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Table 1. General Rubric *(continued)*

Level 2	**Requirements of the task(s) not completed, showing for example;** • only fragmented understanding of the task's mathematical concepts and processes, accompanied by disorganized, incomplete results. • identification of only a few, usually superficial, elements of the task(s). • attempts to address the intended audience that may be incoherent, muddled, or incomplete. • attempts to explain or justify results that are convoluted, illogical, circular, or unrelated to the results shown.
Level 1	**Does not achieve any requirements of the task(s), showing for example;** • an irrelevant, nonsensical, or illegible response that has no valid relationship to the task(s). • no understanding of the task's mathematical concepts and processes. • unsuccessful attempt, if any, to communicate with the intended audience. Usually communication is not attempted. • no attempt to explain or justify results. If attempt is made, it is often unrelated to the task, illegible, or incoherent.

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Score: 6

The Response Shows:

- completion of requirements of the task, including the mathematical processes by which the choice was made.
- comparison of mean, median, and mode, all correctly computed, and the judgment that the differences between these measures for the scores of the two bowlers are not significant.
- two line graphs which indicate clearly the lack of consistency by Bill, which is then chosen as the deciding factor.
- a clear, coherent discussion of these characteristics in a creative and thoughtful way.

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

The tables below show some bowling scores. A higher number indicates a better game.

	D	B
Mean	1457	1473
Median	1425	1245
Mode	141	105

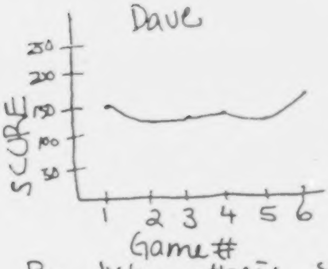
Dave's scores	
152	138
138	141
141	141
144	144
141	152
+ 158	158
874	

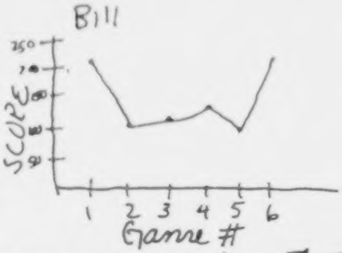
Bill's scores	
210	105
105	105
118	118
131	131
105	210
+ 215	215
884	

Bill has slightly higher total score. Not much difference

Both Dave and Bill are trying out for the bowling team. As the student coach, you must decide which one should be picked to join the team.

Examine the data and use it to justify your choice.
Explain to Dave and Bill how you came to your decision.
Use charts or drawings if it helps you to explain more clearly.





Dave is much more consistent than Bill.

By listing their scores from lowest to highest I found that Dave has a much higher median and mode than Bill.
By adding up Dave's scores and Bill's scores, I found that Bill only has a slightly higher total. There isn't a big difference so I couldn't use that data to choose who should be on the team. By graphing the data, I found out that Dave is much more consistent than Bill, who has many good games yet also many low-scoring games. With these graphs, I can see that both have improved slightly from their first game to their last game. However, I cannot use that to judge them because the difference is slight.

With the data that Dave is much more consistent than Bill, I would chose Dave to be on the team.

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Score: 2

The Response Shows:

- lack of understanding of mean in reporting that the same total of scores and number of games resulted in different "averages."
- incorrect total, which could have been more accurately computed with a calculator.
- a carefully drawn bar graph, but no justification for its inclusion as the graph was not referred to in the explanation.

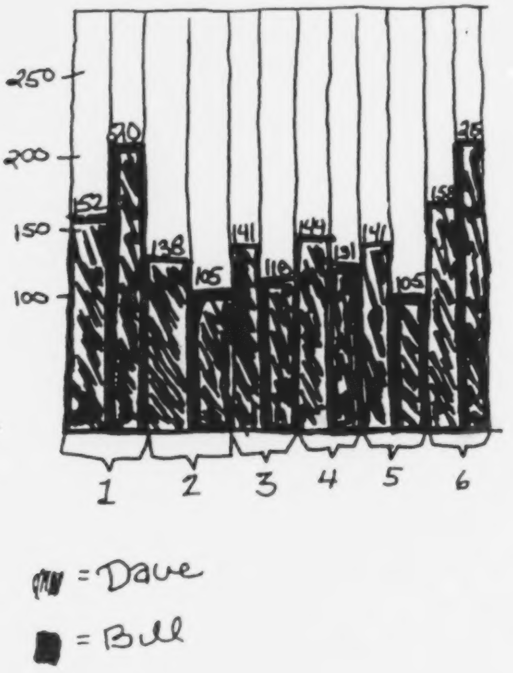
Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

The tables below show some bowling scores. A higher number indicates a better game.

Dave's scores	Bill's scores
152	210
138 290	105 315
141 431	118 433
144 575	131 564
141 716	105 669
158 874	215 874

Both Dave and Bill are trying out for the bowling team. As the student coach, you must decide which one should be picked to join the team.

Examine the data and use it to justify your choice.
Explain to Dave and Bill how you came to your decision.
Use charts or drawings if it helps you to explain more clearly.



Dave's total = 874

Bill's total = 874

Permission for use in training has been granted by the California Department of Education. Materials may be copied for training purposes only.

Sample 5.8

A DAY AT THE CARNIVAL

UTAH DEPARTMENT OF EDUCATION

The Utah Department of Education has developed 30 constructed response items in mathematics for grades 1-6 to complement the multiple-choice tests already in place. These are model assessments, not required assessments. Districts must assess student status with respect to the Utah Core Curriculum goals, but the use of these tests for that purpose is voluntary.

Each of the five scenarios at each grade level contains a series of questions related to the theme of the scenario. The assessments are designed to measure (depending on grade level): logical reasoning, number meanings, number operations, number representation, computation, estimation, algebra, data sets, probability, geometry, measurement, fractions, and patterns. Total assigned points are added up for each scenario. Then points are added up between questions for each of the skill areas being assessed. Four levels of proficiency on each skill are identified (advanced, proficient, basic and below basic) depending on percent of total possible points.

The attached pages are part of one of the sixth grade scenarios.

Dr. Barbara Lawrence
Utah Core Curriculum Performance Assessment Program: Mathematics, 1993
Profiles Corporation
507 Highland Ave.
Iowa City, IA 52240

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

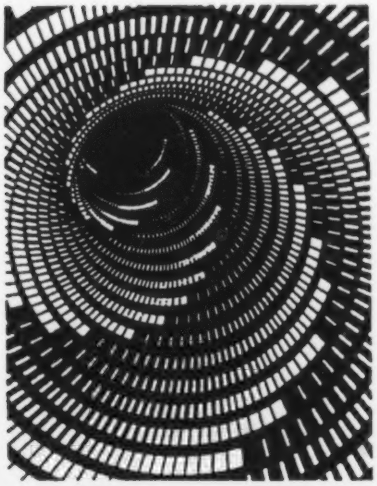
TEACHER DIRECTIONS

Grade 6

MATHEMATICS

A Day at the Carnival

77:M6-B



Utah Core Curriculum Performance Assessment Program

Developed by

PROFILES
CORPORATION

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

ACTIVITY I

Imagine that the carnival is in town. Before you go, you read an article about the carnival in the newspaper.

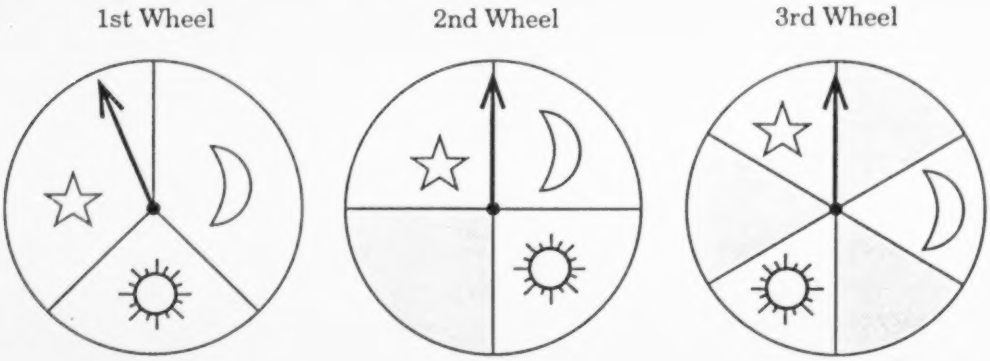
1. In the article, you read that a total of 972 people have attended the carnival in the 3 days that it has been in town and that 1/4 of those people were adults. Based on this information, what was the average number of children that attended the carnival each day? Write your answer in the rectangle. Show your work.

2. As you read further in the article, you discover that the carnival is going to be in town for another 5 days. Using the figures from above, estimate how many children will most likely attend the carnival altogether during the 8 days it is in town. Write your answer in the circle. Show your work.

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

ACTIVITY II

When you arrive at the carnival, the first thing you try is a spinning wheel game. The game uses three different spinning wheels. You need to get the same symbol in each of the three wheels. You get a star when you spin the first wheel, as shown below.



4. What is the probability of getting a star when you spin the second wheel? Write your answer as a fraction.

____

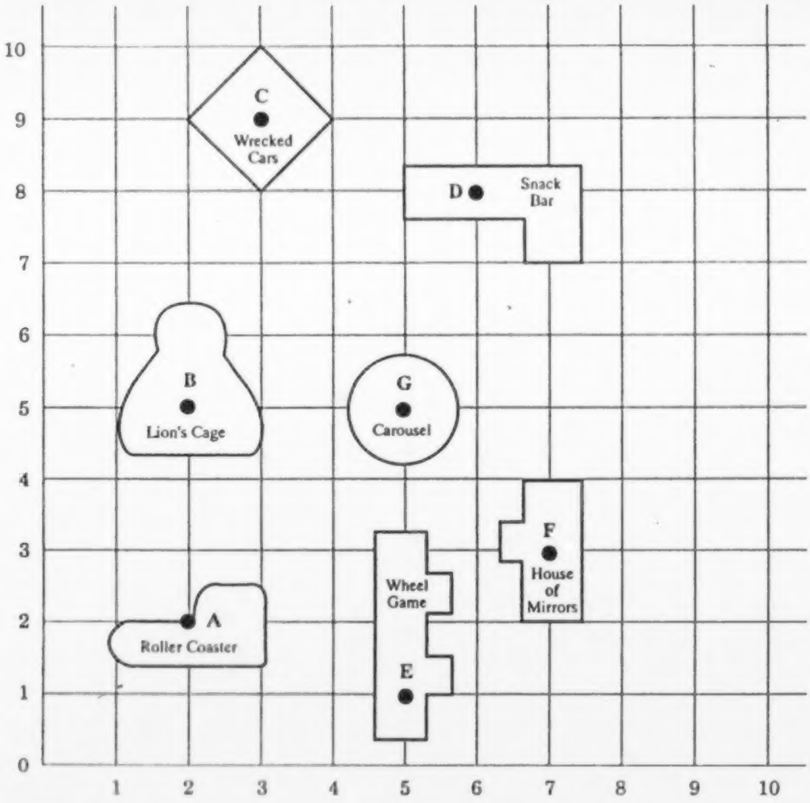
What is the probability of getting a star when you spin the third wheel? Write your answer as a fraction.

____

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

ACTIVITY III

You and your friends want to go to the Snack Bar. You are now near the Roller Coaster (2, 2). On your way to the Snack Bar you want to see three other attractions, so you pull out your map.


C
Wrecked Cars
D
Snack Bar
B
Lion's Cage
G
Carousel
F
House of Mirrors
Wheel Game
E
A
Roller Coaster
0
1
2
3
4
5
6
7
8
9
10

6. Choose a route that leads from the Roller Coaster to three **other** attractions, then to the Snack Bar. Draw your route on the map.

   Now list three ordered pairs of points you will pass through on your way to the Snack Bar.

   ____

   ____

   ____

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

SCORING GUIDE

Introductory Activity Not Scored

Activity I

1. **3 pt.** Student shows the work and the correct result (average number of children–243).

   2 pt. Student addresses the task but something is missing or incorrect; for example, only writes the result.

   1 pt. Student only begins to address the task; for example, shows the work but arrives at an incorrect result.

   0 pt. Student does not answer or answers inappropriately.

2. **3 pt.** Student shows the work and a reasonable estimate (1,920–2,000). (Do not penalize students for using an incorrect number calculated from problem #1.)

   2 pt. Student addresses the task but something is missing or incorrect; for example, only writes the result.

   1 pt. Student only begins to address the task; for example, shows work but arrives at an unreasonable estimate.

   0 pt. Student does not answer or answers inappropriately.

3. **3 pt.** Student decides day and number of persons, calculates how much money is needed; for example, 5 people on Friday, $13.75, and shows work.

   2 pt. Student addresses the task but something is missing or incorrect; for example, selects Thursday but uses Friday rates.

   1 pt. Student only begins to address the task; for example, makes errors in computation.

   0 pt. Student does not answer or answers inappropriately.

Activity II

4. **2 pt.** Student writes 1/4 for the first part and 1/6 for the second part.

   1 pt. Student addresses the task but something is missing or incorrect; for example, answers only one part of the task.

   0 pt. Student does not answer or answers inappropriately.

Permission for use in training has been granted by the Utah State Office of Education. These materials may be duplicated by Utah educators for local classroom use. This permission does not extend to the duplication of materials for commercial purposes.

Sample 5.9

Primary Math Portfolio

Multnomah Educational Service District

The attached pages come from a short handbook designed to give primary grade teachers some ideas on how to begin using portfolios in math for very young students. The author contends that the primary use for portfolios is to assist students to take control of their own learning. Students assemble their portfolios to tell a story about who they are mathematically. Students should be in control of their portfolios, and self-assessment should be emphasized. (The author, however, also points out that there might be other audiences and purposes for the portfolios that might have to be addressed.) The author provides ideas for tasks that students could do to generate material for the portfolio, provides some very practical suggestions for getting started, gives ideas for activities to encourage student self-reflection, discusses student-led conferences, and includes an entire portfolio for a second grade student.

Attached are pages that describe what could go into a portfolio, tips for helping children build portfolios, and draft criteria for assessing the portfolio.

"Portfolio Guidelines in Primary Math," 1994
Leon Paulson
Multnomah County ESD
P.O. Box 301039
Portland, OR 97230
(503) 255-1842

Permission for use in training has been granted by Leon Paulson. Materials may be copied for training purposes only.

What kinds of things can go into a portfolio?

There are many kinds of things suitable for a student's portfolio. Usually, students select items that are part of their classroom curriculum rather than things done specifically for the portfolio. Here are some ideas. As a teacher, your role is to offer students the opportunity to produce things suitable for the portfolio. But it is the student who makes the final selections as a part of telling their story of their learning.

Manipulatives

Encourage students to use rods, blocks, chips, or anything else that helps them solve problems. If the objects are unmanageable, use drawings, photographs, or other ways to document for the portfolio. Encourage younger children to draw pictures of constructions and manipulatives, or use collages to re–create a pattern block or unit block project. Encourage students to talk about the manipulatives as they use them; it will help them to write about it later.

Technology

Encourage students to use computers (printouts are good), calculators, and other devices to demonstrate problem solving. Encourage them to talk and write about what they did.

Group Work

Encourage students to work in groups to solve a problem and encourage them to talk about their ideas. Each child can place evidence of participation in a project in the portfolio along with a description of their personal contribution. If they are not ready to write about it themselves, ask them to dictate their comments to you, an older student, or a volunteer.

Real–world Examples

Encourage students to work on problems that deal with real–world applications. Encourage students to include material completed outside of class! Encourage them to talk and write.

Interdisciplinary Efforts

Use work that is done outside of math class that shows how math is used. Look for portfolio opportunities in science (measurement problems), social studies (graphs; maps that show grids, scale, area, and spatial relationships) or other subjects. Again, encourage them to talk and write.

Journals, Class Publications

Encourage students to look through their journals or books published as class projects to find items that may be appropriate for their portfolios. Have them keep math journals in their portfolios.

Permission for use in training has been granted by Leon Paulson. Materials may be copied for training purposes only.

Tips for helping children build portfolios.

Building a portfolio involves a wide variety of classroom activities. The more variety, the better. It is important to expose students to many kinds of activities so that they have a wide choice of things to put into their portfolios:

- Have students collect all math work in a folder. Periodically (maybe once a week) ask them to look through the folder and select pieces for their portfolio. Have them write why they selected the piece. Accept any reason as valid, but encourage them gradually to become more thoughtful in their reasons. "I liked it," "It shows what I know," "It was fun," "Shows symmetry" are all good reasons. The key is to get them to make a habit of judging their own work.
- Encourage students to talk about how they solve math problems. Form small groups so that they can talk to each other. This will make it easier for them to learn to write about how they solve problems. (Research shows that students encouraged to talk about the problem they're solving do better than those who do not talk.)
- Give children concrete things to talk about and reflect on. Don't ask them to reflect on everything they have done over the year. Ask them to look at two specific examples of their work (say one from early in the year and one from later) and ask them what specific differences they notice. Allow them to discover how much they have learned. Celebrate! Ask them to write about it for their portfolios.
- Encourage students to write about their portfolios as much as possible. Use any special event as an excuse for writing. Parent night, an anticipated visit by the school principal, a visit from another class, a special guest.
- Encourage the students to present their portfolios. A student-led conference (see page 16) is a wonderful opportunity. Put them in charge of setting up and conducting the presentation.

Permission for use in training has been granted by Leon Paulson. Materials may be copied for training purposes only.

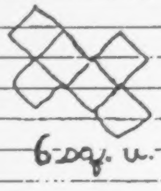
Should students set their own standards?

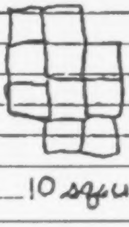
Portfolios offer several opportunities to view standards in a new way. A recent, and exciting development in portfolio assessment is student-negotiated standards. In student-negotiated standards, students take a leading role in setting personal standards that they use to guide their own progress. Here are some ways that you encourage students to set standards.

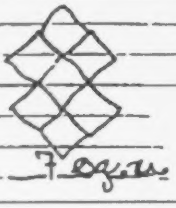
- Distribute some examples of math papers done for another (maybe last year's) class. Ask students to look at the papers and decide which are the better papers. Have students brainstorm a list of characteristics of good papers. Encourage them to figure out, to *discover* what constitutes good math work. Then have each student select one or more characteristic to work on in their own papers. Help them build a rubric or checklist to keep track of their own progress.
- Teach the students to assess work using a set of process guidelines such as those appearing on page 18. Don't worry that their "scores" agree with yours, but do work for them to make consistent judgments so that they can see their own growth. Encourage students to talk to you and to each other about what constitutes good math problem solving — help them discover that doing math well involves much more than merely finding the right answer.
- Have students interview adults, especially adults with technical training to find out what real mathematicians do. Have the class build a list of things mathematicians do. Have students look for examples in their portfolios that correspond to the items on this list.

This is a relatively new area and there are few guidelines in education. Negotiated standards have been successful in industry, especially in the work of W. Edwards Deming. For an example of student-negotiated standards in a writing portfolio, see Paulson, Paulson & Frazier (in press).

Now I know how to find the area in a shape. Here are some samples:







Permission for use in training has been granted by Leon Paulson. Materials may be copied for training purposes only.

A Guide for Judging Portfolios
by Leon and Pearl Paulson

An **Outstanding** Portfolio. An outstanding portfolio is a coherent story of the student as a reflective learner where all the parts of the portfolio bear a clear relationship to each other and to a central purpose. There is an awareness of the perspectives of other stakeholders, and the student's self-assessment has been enhanced by this knowledge. A reviewer can look at the portfolio and easily understand how the judgments about the learner came to be made and the degree to which different stakeholders would agree. When reviewing the portfolio, outsiders get the feeling they really know the person whose achievement is depicted there, and have a fair understanding of how the learning came about.

An **On-Track** Portfolio. An on-track portfolio is in the process of becoming a story of the student as an independent learner. There are relationships between one part of the portfolio and another. There is evidence of student ownership. The learner has a personal investment in selecting and explaining the content. It is possible to distinguish other stakeholders' goals from the student's or to recognize instances when they overlap. The portfolio may be created for others to assess, but there is also evidence of self-assessment. The student's voice is always audible.

An **Emerging** Portfolio. In an emerging portfolio there is a sense of intentionality controlling some of the student's choices. Students may not be able to verbalize the reasons, even as they reflect on their choices, but the reviewer may be able to recognize a relationship between some exhibits or infer the reasons. Or, there may be evidence that the student had some insight into the teacher's purposes. While evidence of self-reflection adds information to the presentation, at this point in the development of the portfolio there is insufficient information or organization to characterize the portfolio as either a story of learning or a portrait of the learner.

An **Off-track** portfolio. An off-track portfolio is simply a container of student work or assessments, without an attempt on the part of the learner to provide organization. There is no attempt by the learner to make a coherent statement about what learning has taken place. The child's understanding of the task is minimal—the portfolio is about "collecting what the teacher asks for." For the student, the portfolio was built by following instructions. Self-reflective statements, if present, add little to clarify organization or explain learning.

Based on "A Guide for Judging Portfolios" by F. L. and Pearl R. Paulson. (Available from the author.)

Permission for use in training has been granted by Leon Paulson. Materials may be copied for training purposes only.

Sample 6.1

Hawaii Algebra Grading Process

Hawaii Algebra Learning Project, University of Hawaii

The Hawaii Algebra Learning Project is a research-based curriculum and assessment package developed at the University of Hawaii. The assessment component focuses heavily on communication and problem-solving skills as applied to key concepts.

Attached are several pages related to how students are graded. First, students write regularly in journals (roughly 40 percent of the final grade). The journal prompts for Chapter 2 are shown. Following the journal prompts is a sample student response with the rubric used to score all journal entries. Next is a scoring sheet for the group presentations of problems solved (roughly 40 percent of the grade). Finally, a sample student record sheet for a chapter is shown. Students receive a "discussion" score on the problem-solving tasks and a journal score ("JW") for each journal prompt.

Annette Matsumoto, Project Director
The Hawaii Algebra Learning Project
Curriculum Research and Development Group
UHS3-227
1776 University Avenue
Honolulu, HI 96822

Permission for use in training has been granted by the Hawaii Algebra Learning Project. Materials may be copied for training purposes only.

NAME: ______________________________

CHAPTER 2

Problem Set	JOURNAL PROMPT	Completed
2–1	Describe how a third-grader might use the commutative property.	
2–2	Is division a commutative operation? Why or why not?	
2–3	Manfred the Martian is teaching you a new math. He uses the symbol ⌂. Tell what it might mean and give some examples. Is it commutative?	
2–4	Describe your method for determining whether a number is greater than or less than another number.	
2–5	Why do we need negative numbers?	
2–7	The pebble model helps/does not help me solve problems because . . .	
2–8	Write a story problem or a word problem that can be solved by $-7 - (-8) = 1$.	
2–9	Describe what helps you remember the order of operations.	
2–10	Order of operations is needed because . . .	
2–11	When adding two negative numbers, the sum is negative. When multiplying two negative numbers, the product is positive. Why is this true?	
2–12		

Hawaii Algebra Learning Project, University of Hawaii

Permission for use in training has been granted by the Hawaii Algebra Learning Project. Materials may be copied for training purposes only.

Scoring for Journal Prompts

Criteria for Journal Scoring

4 = a. Prompt is completed.
b. Support is given for the prompt completion by using either examples or counter examples.
c. Ideas are communicated to the reader.
d. Legible/readable
e. Questions are suggested for continued reflection or extensions are given.

3 = Omission of one criteria from level 4.

2 = Omission of two criteria from level 4.

1 = Omission of three criteria from level 4.

0 = Omission of more than three criteria from level 4.

Journal Prompt

"A good problem is..."

Sample 8th Grade Student Response:

A good problem is one that lets me do anything I want. I like to play with the word problems. Sometimes you can change the problem to make it simpler or harder and then you have a different one to solve. I never thought I would rather do a word problem than one like 5 + 7 + 8.

My favorite word problem is like the kissing cousins problem. You can solve it with a diagram, a picture, a list, or a formula. You can make the problem harder by changing the number of cousins. Like it could be 15 cousins instead of 5. If you really wanted to make it hard, you could change it to 100 or more.

Hawaii Algebra Learning Project, University of Hawaii

Permission for use in training has been granted by the Hawaii Algebra Learning Project. Materials may be copied for training purposes only.

Scoring Sheet for Group Problem-Solving Presentations

Problem Set #: ____________________
Group: ____________________
Presenter: ____________________

Criteria	Good	Fair	Needs Improvement
Explained clearly.			
Explained thinking, not just steps.			
Asked for other solutions from the class.			
Presented more than one solution, if available from the group.			
Extended problem by presenting to the class a new problem made from the presented one or by noticing patterns in the problem or by noting similarities between this problem and one previously done.			
Asked class good questions, like is there another way?, is the answer unique?, what if...			
Answered class questions.			
Neat transparency with group member names on it.			
Spoke loudly and clearly.			
If answer was incorrect, used it to create a discussion.			
Group is not rude to others in or out of their group.			
Group is supportive by being attentive when someone from the group is presenting or contributing.			
Group participated in other problem discussion.			
Group worked together.			
All members of group had homework.			
Listened to other people's ideas on homework or correcting solutions and answers.			

★ Students developed these criteria.

Permission for use in training has been granted by the Hawaii Algebra Learning Project. Materials may be copied for training purposes only.

Points possible: 4

Points received: _______

Comments: (If you are not happy with your group grade, you may individually or in your group redo your discussion on paper and turn it in as a revision.)

Hawaii Algebra Learning Project, University of Hawaii

Permission for use in training has been granted by the Hawaii Algebra Learning Project. Materials may be copied for training purposes only.

Student Record Sheet–Chapter 4

Activity Chapter 4	4	3	2	1	0
Discussion LB/LA					
JW LB/LA					
Discussion 4-1	✓				
JW 4-1		✓			
Discussion 4-2			✓		
JW 4-2		✓			
Discussion 4-3					✓
JW 4-3				✓	
Discussion 4-4		✓			
JW 4-4			✓		
Discussion 4-5		✓			
JW 4-5		✓			
Quiz				✓	
Discussion 4-6			✓		
JW 4-6		✓			
Discussion 4-7					
JW 4-7					
Discussion 4-8					
JW 4-8					
Discussion 4-9					
JW 4-9					
Discussion 4-10					
JW 4-10					
Chapter 4 Test					
LAB ACTIVITY I					
LAB ACTIVITY II					

•First Name• •Last Name• Chapter 4

Permission for use in training has been granted by the Hawaii Algebra Learning Project. Materials may be copied for training purposes only.

Appendix B

Resources *

Contents: **Page**

Science and Mathematics Reform Initiatives 2

Agencies 6

State Departments of Education and School Districts 11

Publishers 14

Print Materials 15

- NCME Instructional Modules 15
- Sources of Information on Tests 17
- Articles and Books 18
- Annotated Bibliographies 23
- Videotape Resources 25

---

*Some listings based on information in Patricia Wheeler and Geneva Haertel, *Resource Handbook on Performance Assessment and Measurement*, Owl Press, PO Box 89, Berkeley, CA 94701, 1993.

The *Toolkit* and Reform in Mathematics and Science Education

In response to publicized reports that U.S. students were lagging behind in international comparisons of academic performance, along with projections of changing knowledge and skills needed to compete in a global workforce, numerous efforts at reforming mathematics and science education have arisen. A brief list of the better known efforts is included in the references at the end of this chapter.

While the list is by no means exhaustive (it should definitely be supplemented by local site-based management efforts), it does provide an indication of the scope of the reform effort in mathematics and science education.

Research in education and psychology support a changing view of how learning occurs (Brooks and Brooks, 1993; Davis and Maher, 1990; Lorsback and Tobin, 1992; Wheatley, 1991). In this new perspective, the learner actively constructs personal meaning from information and experiences by linking the new information with pre-existing knowledge. (From this perspective, traditional modes of instruction and assessment often fall short, resulting in a de-emphasis on individual student learning, understanding, and mastery of basic process and problem-solving skills.

This "constructive" learning process requires more reflection, collaboration, and facilitation. Students must be given opportunities to rearrange, add to, or delete information from [their] personal conceptual framework[s] (p. 14). Since assessment is intimately linked with instruction, it too must change. The assessment approaches teachers select send a very clear signal to the student about what is important. Questions must be asked about what students should know and be able to do. Alternative forms of assessment are needed if the intention is to assess individual understanding and attainment of process and critical-thinking skills or to acknowledge the value of each student.

How can the *Toolkit* assist in this reform effort? The *Toolkit* has been designed specifically as a professional development resource. It is a compilation of activities and supportive materials developed not only as an alternative assessment resource, but as a means of engaging teachers in dialog about changing the way mathematics and science have traditionally been taught and assessed.

The movement to establish curriculum standards that specify what a student should know and be able to do is an attempt to spell out what it is that we, as a country, value as outcomes of the U.S. educational system. By getting involved in the articulation of these outcomes and participating in discussions on how best to asses the extent to which they have been achieved, teachers can become active agents in the change process.

Science and Mathematics Reform Initiatives

Selected organizations, programs and publications that have had a major impact on recent mathematics and science education reform. See also *Guide to Reform in Mathematics and Science Education: the Big Players*, 1994.

- **Project 2061**, funded by the American Association for the Advancement of Science (whose name signifies the next year comet Halley will be visible from earth), consists of three phases. The first resulted in *Science for All Americans* (1989), which focuses on science literacy. The second resulted in *Benchmarks for Science Literacy* (1993), which includes "statements of what all students should know or be able to do in science, mathematics, and technology by the end of grades 2, 5, 8, and 12" (p. xi). The final phase will include the development of implementation strategies.

- **The Mathematical Sciences Education Board** is composed of professional mathematics education associations and reports directly to the National Research Council. It has published several reports pertaining to mathematics reform: *Everybody Counts* (1989), *Reshaping School Mathematics* (1990), and *Measuring Up* (1993).

- The **New Standards Project** is a "grassroots partnership" of 19 states and 6 urban school districts "that is adopting a set of very high national education standards and developing a new kind of assessment system designed to gauge student progress toward those standards" (see *The New Standard*, 1994, p. 4)

- The movement to specify national standards (statements of what students should know [content] and be able to do [performance] upon completion of each grade level) was spearheaded by the **National Council of Teachers of Mathematics** with the publications of *Curriculum and Evaluation Standards* (1989), *Professional Standards for Teaching Mathematics* (1991), and *Assessment Standards for School Mathematics* (Working Draft, October 1993). Professional organizations in other subject areas have followed suit, with the result that the National Education Goals Panel adopted guidelines for national standards, stressing that they should be voluntary, academic, world-class, bottom-up, useful, and adaptable ("Goals," 1993). National science standards are being developed by the National Research Council's **National Committee on Science Education Standards and Assessment**. Preliminary drafts have been circulated for comments (*National Science*, 1993), with final versions scheduled to be published in the fall of 1994. Additional non-curricular standards are also being developed for school/system delivery, opportunity to learn, and professional licensing of teachers.

- The National Science Foundation, in addition to supporting numerous research and professional development projects in mathematics and science education, also supports three programs designed to foster systemic change: The **Statewide Systemic Initiative** provides funds to states for long-term mathematics and science education projects demonstrating collaboration between the governor, chief state school officers, teacher organizations, legislature, corporations, and presidents of colleges and universities. **Urban Systemic Initiatives** and **Rural Systemic Initiatives** must demonstrate similar collaboration among a range of different constituencies involved in mathematics and science education. All three projects must show a commitment to reforming science and mathematics education in a systemic way. **Scope, Sequence and Coordination**, funded by the National Science Foundation, is a project that advocates "presenting key science concepts, appropriately sequenced, manageable in their scope, and coordinated within and between the science disciplines." Begun in the middle school and now expanding to high school, the project has published *The Content Core: A Guide for Curriculum Developers* (1992, p. 2).

- The Eisenhower National Program, funded by the U.S. Department of Education, provides **Title II** funds to support professional development for mathematics and science teachers and funds the development and implementation of statewide systemic reform efforts through its **State Curriculum Frameworks for Mathematics and Science Education Program**. The Eisenhower Program also supports ten **Regional Consortia for Mathematics and Science Education**, each of which offers technical assistance to support state, regional, and local reform initiatives. A **National Clearinghouse** was also established in 1993 to serve as a center for information on mathematics and science education reform. It is developing a database of instructional materials, methods, and assessment resources.

- The U.S. Department of Education also funds ten **Regional Education Laboratories** that identify regional education needs and develop programs to meet those needs. The ten laboratories have collaborated as **The Laboratory Network Program** to produce this *Toolkit* and *A Database and Catalog of Alternative Assessments* (1994) focusing on alternative assessment; a companion document, *Facilitating Systemic Change in Science and Mathematics Education: A Toolkit for Professional Developers;* and two documents focusing on curriculum frameworks, *Curriculum Framework Analysis Tool* (1992) and *A Summary of Analyzed State Curriculum Frameworks* (1993).

- **The Goals 2000: Educate America Act** was passed in March 1994. Several of its eight provisions relate directly to mathematics and science education reform:

  — Goal 3: American students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, math, science, arts, foreign languages, history and geography, civics and government, and economics.

  — Goal 4: The nation's teaching force will have access to programs for the continued improvement of their professional skills and the opportunity to acquire the knowledge and skills needed to instruct and prepare all American students for the next century.

  — Goal 5: U.S. students will be first in the world in math and science achievement.

  — Goal 6: Every American will be literate and will possess the knowledge and skills necessary to compete in a global economy.

References

Assessment Standards for School Mathematics. (Working Draft, October 1993). Reston, VA: National Council of Teachers of Mathematics.

Benchmarks for Science Literacy. (1993). New York: Oxford University Press.

Brooks, J. G. and M. G. Brooks. (1993). *In Search of Understanding: The Case for Constructivist Classrooms.* Alexandria, VA: Association for the Supervision of Curriculum and Development.

The Content Core: A Guide for Curriculum Developers. (1992). Washington, DC: The National Science Teachers Association.

Curriculum and Evaluation Standards for School Mathematics. (1989). Reston, VA: National Council of Teachers of Mathematics.

Curriculum Framework Analysis Tool. (1992). Aurora, CO: Mid-continent Regional Educational Laboratory.

Everybody Counts: A Report to the Nation on the Future of Mathematics Education, (1989). Washington, DC: Mathematics Sciences Education Board.

Facilitating Systemic Change: A Toolkit for Professional Developers. (Pilot Test Version, 1993). Andover, MA: The Regional Laboratory for Educational Improvement of the Northeast and Islands.

"Goals Panel Adopts Principles on National Content Standards," *Education Daily* 126 (122): 3. November 17, 1993.

Guide to Reform in Mathematics and Science Education: The Big Players. (Spring 1994). Austin, TX: Southwest Educational Development Laboratory.

"Learner-Centered Psychological Principles, Guidelines for School Redesign and Reform." (January 1993). Aurora, CO: Mid-continent Regional Educational Laboratory.

Lorsback, A. and K. Tobin. (September 1992). "Constructivism as a Referent for Science Teaching." *Journal for Research in Science Teaching* (30).

Maher, C. A. and R. B. Davis. (1990). "Building Representations of Children's Meanings." In *Journal for Research in Mathematics: Constructivist Views on the Teaching and Learning of Mathematics.* Reston, VA: National Council of Teachers of Mathematics.

Measuring Up: Prototypes for Mathematics Assessment. (1993). Washington, DC: Mathematics Sciences Education Board.

National Science Education Standards: An Enhanced Sampler. (February 1993). Washington, DC: National Research Council.

The New Standard. (May 1994). Washington, DC: New Standards Project.

Professional Standards for Teaching Mathematics. (1991). Reston, VA: National Council of Teachers of Mathematics.

Regional Educational Laboratory Network Program. (1994) *Improving Science and Mathematics Education—A Database and Catalog of Alternative Assessments.* Portland, OR: Northwest Regional Educational Laboratory.

________. (1994) *Improving Science and Mathematics Education—A Toolkit for Professional Developers: Alternative Assessment.* Portland, OR: Northwest Regional Educational Laboratory.

Reshaping School Mathematics: A Philosophy and Framework for Curriculum. (1990). Washington, DC: Mathematics Sciences Education Board.

Science for All Americans. (1989). Washington, DC: American Association for the Advancement of Science.

Spector, B. S. (1993). "Order Out of Chaos: Restructuring Schooling to Reflect Society's Paradigm Shift." *School Science and Mathematics,* 93(1): 9-19.

A Summary of Analyzed State Curriculum Frameworks. (1993). Aurora, CO: Mid-continent Regional Educational Laboratory.

Update: Project 2061, Education for a Changing Future. (1992). Washington, DC: American Association for the Advancement of Science.

Wheatley, G. H. (1991). "Constructivist Perspectives on Science and Mathematics Learning." *Science Education* 75(1): 9-21.

AGENCIES

The agencies on this list are all involved in alternative assessment at some level. If no specific description is given of the types of activities in which an agency is engaged, it is very likely that they are doing one or more of the following: developing content standards, developing alternative assessments, developing materials to assist others in the development of alternative assessment, or conducting training on alternative assessment. Including agencies and publishers on this list should not be taken as an endorsement of their services or products.

American Association for the Advancement of Science (AAAS), 1333 H Street NW, Washington, D.C. 20005.

Has information on alternative assessment, promising practices, and curriculum standards in science.

Annenberg/CPB Math and Science Project

"The Guide to Math & Science Reform" is a database of reform initiatives, organization, materials, and events.

American Educational Research Association (AERA), 1230 Seventeenth St., NW, Washington, DC 20036, (202) 223-9485.

Has a subcommittee working on assessment standards.

Appalachia Educational Laboratory, 1031 Quarrier St., PO Box 1348, Charleston, WV 25325, (304) 347-0400, Helen Saunders.

Training and materials on alternative assessment.

Assessment Training Institute, 215 SW Washington, Suite 201, Portland, OR 97204, (503) 228-3060. Rick Stiggins

Training, videos, print material on classroom assessment.

Association for Supervision and Curriculum Development (ASCD), 1250 N. Pitt St., Alexandria, VA 22314, (703) 549-9110.

Coordinates consortium efforts in assessment development, and has publishes print and video resources. Contact Kathy Busick (PREL) for the "Authentic Assessment Network" or Shelly Lipowich (602-299-9583) for a project on assessing math and science process skills.

Buros Institute of Mental Measurements, The University of Nebraska—Lincoln, PO Box 880348, 135 Bancroft Hall, Lincoln, NE 68588, (402) 472-6203.

Conducts and publishes reviews of tests.

Center for Civic Education, 5146 Douglas Fir Rd., Calabasas. CA 91302.

Coordinating the effort to develop curriculum standards in civics.

Center for the Study of Reading, 174 Children's Research Center, 51 Gerty Dr., Champaign, IL 61801.

Coordinating the development of curriculum standards in English.

Center on Learning, Assessment and School Structure (CLASS), 39 Main St., Geneseo, NY 14454, Grant Wiggins.

Training and videos on assessment.

Coalition of Essential Schools, Brown University, Box 1969, Providence, RI 02912, (401) 863-3384, Joe McDonald.

Coordinates cooperative efforts on alternative assessment; disseminates *Exhibitions Collection*, a sampler of performance assessments.

Council of Chief State School Officers (CCSSO), 1 Massachusetts Ave., NW, Suite 700, Washington, DC 20001, Ed Roeber (202) 336-7045 or Linda Bond (371) 328-0639.

Conducts yearly surveys of state assessment activities.

Council on Educational Development and Research (CEDaR), 2000 L St., NW, Suite 601, Washington, DC 20036, (202) 223-1593.

Coordinates the nation's Regional Laboratories and Research Centers, several of which are involved in alternative assessment projects.

Educational Testing Service (ETS), Rosedale Rd., Princeton, NJ 08541, (609) 921-9000.

Training and materials on assessment; develops assessments under contract.

ERIC Clearinghouse on Assessment and Evaluation, Catholic University of America, 209 O'Boyle Hall, Washington, DC 20064, (202) 319-5120.

Library of articles and information on assessment. The collection is available on-line.

Far West Laboratory for Educational Research and Development (FWL), 730 Harrison St., San Francisco, CA 94107, (415) 565-3000.

Training and materials on assessment; develops assessments under contract.

Laboratory Network Program (LNP), see attached contact list for phone numbers.

Science and math alternative assessment database and training materials. Database is also available on Internet. Some Labs have lending libraries.

Lawrence Hall of Science, University of California, Berkeley, CA 94720.

Publishes instructional materials and assessment instruments in science.

Mid-continent Regional Educational Laboratory (McREL), 2550 S. Parker Rd. Suite 500, Aurora, CO 80014, (303) 337-0990.

Music Educators National Conference, 1902 Association Dr., Reston, VA 22091.

Coordinating development of curriculum standards in the arts.

National Assessment of Educational Progress, (NAEP), Educational Testing Service, Princeton, NJ 08541, (609) 921-9000.

Conducts national assessments of student achievement in reading, writing, mathematics, science, fine arts, and social studies. Some components are performance based.

National Assessment Governing Board (NAGB), 800 N. Capitol St. NW, Suite 825, Washington, DC 20002, (202) 357-6938.

Is the governing body for the *National Assessment of Educational Progress*.

National Association for Sports and Physical Education, 1900 Association Dr., Reston, VA 22091.

Coordinating development of curriculum standards in physical education.

National Center for History in the Schools, UCLA, 231 Moore Hall, 405 Hilgard Ave., Lost Angeles, CA 90024.

Coordinating development of curriculum standards in history.

National Center for Research in Mathematical Sciences Education, University of Wisconsin, 1025 W. Johnson St., Madison, WI 53706, Don Chambers.

National Center for Research on Evaluation, Standards and Student Testing (CRESST), UCLA Graduate School of Education, 145 Moore Hall, 405 Hilgard Ave., Los Angeles, CA 90024, (310) 206-1532, Ron Dietal.

Internet database on alternative assessment in all subject areas; training and research on alternative assessment.

National Center for the Study of Writing, University of California at Berkeley, 5513 Tolman Hall, Berkeley, CA 94720, (510) 643-7022.

National Council of Geographic Education, 1600 M St., N.W., Washington, DC 20036.

Coordinating development of curriculum standards in geography.

551

National Council on Measurement in Education (NCME), 1230 Seventeenth St., NW, Washington, DC 20036, (202) 223-9485.

Has a working committee on assessment standards.

National Council for the Social Studies, 3501 Newark St., N.W., Washington, DC 20016.

Coordinating development of curriculum standards in social studies.

National Council of Teachers of English (NCTE), 1111 Kenyon Road, Urbana, IL 61801, (800) 369-NCTE.

Regularly publishes information on alternative assessment in reading, writing, speaking and listening.

National Council of Teachers of Mathematics (NCTM), 1906 Association Dr., Reston, VA 22091, (800) 235-7566.

Regularly publishes information on alternative assessment and curriculum standards in mathematics.

National Science Education Standards, 2101 Constitution Ave., N.W., HA 486, Washington DC 20418.

Coordinating development of curriculum standards in science.

New Standards Project, University of Pittsburgh, Learning Research and Development Center, 3939 O'Hara St., Pittsburgh, PA 15213, (412) 624-8319.

Coordinates cooperative efforts in alternative assessment.

North Central Regional Educational Laboratory (NCREL), 1900 Spring Rd., Suite 300, Oak Brook, IL 60521, (708) 571-4700, Jeri Nowakowski.

Publishes assessment information including a database on state assessment activities and monographs on legal issues.

Northwest Evaluation Association (NWEA), 5 Centerpointe Dr., Suite 100, Lake Oswego, OR 97035, (503) 624-1951, Alan Olson.

Training in assessment, computer-adaptive testing.

Northwest Regional Educational Laboratory (NWREL), Test Center, 101 SW Main St., Suite 500, Portland, OR 97204, (503) 275-9582.

Lending library of alternative assessment instruments and information, bibliographies, and training on assessment.

Office of Educational Research and Improvement (OERI), 555 New Jersey Ave., NW, Washington, DC, (202) 219-1828.

Funds many assessment projects.

Pacific Region Educational Laboratory (PREL), 828 Fort Street Mall, Suite 500, Honolulu, HI 96813, (808) 533-6000, Kathy Busick.

Professional development and materials on classroom and large-scale assessment, performance standards, cultural considerations in assessment.

Performance Assessment Collaborative for Education (PACE). Harvard University. Dennie Palmer Wolf.

Ongoing, long-term projects in support of quality assessment; institutes, research, technical assistance, consulting.

Portfolio Assessment Clearinghouse, San Dieguito Union High School, 710 Encinitas Blvd., Encinitas, CA 92024.

Newsletter on portfolio projects.

Regional Laboratory for Educational Improvement of the Northeast and Islands, The, 300 Brickstone Square, Suite 950, Andover, MA 01810

Publishes "Facilitating Systemic Change in Science and Mathematics Education: A Toolkit for Professional Developers," collaboratively developed by the 10 regional educational laboratories.

Southeastern Regional Vision for Education, PO Box 5367, Greensboro, NC 27435, (800) 755-3277.

Publishes documents on assessing, including "How to Assess Student Performance in Science, 1993."

Southern California Alliance for Physics Improvement, Dr. Roger Nanes, California State University, Fullerton, (714) 733-2188 or Dr. John Jewett, California State University, Pomona,
(909) 869-4022.

Project SCAMPI is a multi-year project for the development of mentor teachers in physics. Assessment is the focus area.

STATE DEPARTMENTS OF EDUCATION AND SCHOOL DISTRICTS

State departments of education and school districts are in the forefront of developing alternative assessments. Presence on this list does not necessarily imply endorsement of the assessments being developed.

Alberta Education, Box 43, 11160 jasper Ave., Edmonton, AB T5K 0L2, Canada.

Alberta has developed performance assessments for ninth graders in science and mathematics. They have also development handbooks to assist teachers with classroom assessment.

Arizona Department of Education, 1535 W. Jefferson, Phoenix, AZ 85007, (602) 542-5754, Paul Koehler.

California State Department of Education, 721 Capitol Mall, Sacramento, CA 95814, (916) 657-3011, Dale Carlson.

The *California Learning Assessment System* (CLAS) is a series of multiple-choice, short answer and performance assessments for grades 4, 8, and 10 in social studies, science writing, mathematics and reading. CLAS is designed to track the progress of individual students. The *Golden State Exam* (GSE) and the *Career-Technical Assessment Program* (C-TAP) are voluntary high school tests which have performance components.

Connecticut Department of Education, Division of Research, Evaluation and Assessment, Box 2219, Hartford, CT 06145, (203) 566-1684, Douglas Rindone.

Connecticut has developed performance assessments in writing, science, math, reading and social studies.

Illinois State Board of Education, 100 N. 1st St., Springfield, IL 62777.

Illinois was one of the first states to redesign its statewide reading assessment. They also have a direct assessment in writing.

Kansas State Board of Education, 120 S.E. 10th Ave., Topeka, KS 66612.

Kansas has developed a pilot open-ended math assessment for grades 4, 7, and 10.

Kentucky Department of Education, Office of Assessment and Accountability, 19th Floor Capital Plaza Tower, 500 Mero St., Frankfort, KY 40601, (502) 564-4394, Edward Reidy.

The *Kentucky Instructional Results Information System* (KIRIS) is developing short response items in reading, math, science, and social studies.

Maryland Department of Education, 200 W. Baltimore St., Baltimore, MD 21201, (410) 333-2000, Steven Ferrara.

The *Maryland School Performance Assessment Program* (MSPAP) is a series of assessment tasks for grades 3, 5, and 8 which are integrated across reading, writing, science, math, and social studies. These are used for school accountability.

Maryland Assessment Consortium, c/o Frederick County Public Schools, 115 E. Church St., Frederick, MD 21701, (301) 694-1337. Jay McTighe

Cooperative development of performance assessments in science, math, reading, writing, and social studies; video and other training materials.

Massachusetts Department of Education, 1385 Hancock St., Quincy, MA 02169, 617-770-7334.

The *Massachusetts Educational Assessment Program* (MEAP) has developed open-ended assessments in reading, social studies, science and math for grades 4, 8 and 12.

Michigan Department of Education

Michigan designed an *Employability Skills Portfolio* to assist students to demonstrate that they have acquired the skills needed for employment.

New York State Education Department, Division of Educational Testing, Room 770 EBA, Albany, NY 12234, (518) 474-5902, Carolyn D. Byrne.

Oregon Department of Education, Public Service Building, 255 Capitol St., NE, Salem, OR. (503) 378-8004, Michael Dalton.

Oregon has conducted writing assessments using a six-trait model since 1985 and is developing on-demand performance assessment, projects and portfolios in grades 3, 5, 8, and 10 in science, reading, and math.

Pittsburgh Public Schools (ARTS PROPEL), Division of Writing & Speaking, 341 S. Bellefield Ave., Pittsburgh, PA 15213, FAX: (412) 622-8578.

Arts PROPEL has developed portfolios in creative writing, music, and art.

Texas Education Agency, 1701 North Congress Ave., Austin, TX 78701, (512) 463-9536, Barbara Green.

Texas has developed performance assessments for social studies and science in grades 4 & 8.

Vermont Department of Education, 120 State Street, Montpelier, VT 05620, (802) 828-3135, Richard P. Mills, Commissioner

Vermont is developing portfolios for grades 4, and 8 in writing and math problem-solving. The purpose is program evaluation and improvement.

Waialae Elementary School, 1045 19th St., Honolulu, HI 96816, (808) 737-4880.

Waialae is developing a multi-dimensional assessment system using portfolios and other performance assessments.

PUBLISHERS

Many test publishers have performance assessment components available for their achievement test series. These range from short, fill-in-the-blank type questions to more in-depth, integrated tasks. Norms are frequently available. Presence on this list does not necessarily imply endorsement of products.

American Guidance Service, 4201 Woodland Road, P.O. Box 99, Circle Pines, MN 55014, (800) 328-2560.

CTB/McGraw-Hill, PO Box 150, Monterey, CA 93942, (800) 538-5380.

Publishes standardized multiple-choice tests (CTBS, CAT) and related performance assessments.

Educational Testing Service (ETS)/National Assessment of Educational Progress (NAEP), Educational Testing Service, Princeton, NJ 08541, (800) 223-0267.

The NAEP is administered yearly on a national sample of students. Some assessments have performance components. ETS is also involved in a number of other assessment projects.

IOX Associates, 5301 Beethoven St., Suite 109, Los Angeles, CA 90066, (310) 822-3275.

Publishes training videos on assessment.

Psychological Corporation, 555 Academic Court, San Antonio, TX 78204, (800) 228-0752.

Publishes standardized multiple-choice tests (SAT, MAT) and related performance assessments (*GOALS*—short answer; *Integrated Assessment System*—longer, more integrated tasks).

Riverside Publishing Company, 8420 Bryn Mawr Ave., Chicago, IL 60631, (800) 323-9540.

Publishes standardized multiple-choice tests (ITBS, TAP, ITED) and related performance assessments.

Touchstone Applied Science Associates (TASA), Fields Lane, P.O. Box 382, Brewster, NY 10509, (914) 277-4900.

Publishes *Degrees of Reading Power*, which uses cloze questions.

NCME Instructional Modules

Since 1987, the National Council on Measurement in Education (NCME) has provided a series of instructional units, called ITEMS, in its quarterly publication, *Educational Measurement: Issues and Practice*. The list of ITEMS through the fall of 1993 is provided below:

Arter, Judith A. and Vicki Spandel. (1992, Spring). Using portfolios of student work in instruction and assessment. *11*(1), 36-44.

Brennan, Robert L. (1992, Winter). Generalizability theory. *11*(4), 27-34.

Brookhart, Susan. (1993, Spring). Assessing student achievement with term papers and written reports. *12*(1), 40-47.

Cook, Linda L. and Daniel R. Eignor. (1991, Fall). IRT equating models. *10*(3), 37-45.

Frary, Robert R. (1988, summer). Formula scoring of multiple choice tests (Correction for guessing). *7*(2), 33-38.

Frisbie, David A. (1988, Spring). Reliability of scores from teacher-made tests. *7*(1), 25-35.

Frisbie, David A. and Kristie K. Waltman. (1992, Fall). Developing a personal grading plan. *11*(3), 35-42.

Hambleton, Ronald K. and Russell W. Jones. (1993, Fall). comparison of classical test theory and item response theory. *12*(3), 38-47.

Harris, Debrah. (1989, Spring). Comparison of 1-, 2-, and 3-parameter IRT models. *8*(1), 35-41.

Harvill, Leo M. (1991, Summer). Standard error of measurement. *10*(2), 33-41.

Kolen, Michael J. (1988, Winder). Traditional equating methodology. *7*(4), 29-36.

Oosterhof, Albert C. (1987, Winter). Obtaining intended weights when combining student scores. *6*(4), 29-37.

Osterlind, Steven. (1988, Fall). Using CRTs in program curriculum evaluation. *7*(3), 23-30.

Stiggins, Richard J. (1987, Fall). Design and development of performance assessments. *6*(3), 33-42.

Stiggins, Richard J. (1992, Summer). High quality classroom assessment: What does it really mean? *11*(2), 35-39.

Traub, Ross E. and Glenn L. Rowley. (1991, Spring) Understanding reliability. *10*(1), 37-45.

Sources of Information on Tests

Educational Testing Service. *The ETS Test Collection Catalog.* Phoenix, AZ: Oryx Press, 1989-92.

Volume 1: Achievement Tests (2nd ed.), 1992
Volume 2: Vocational Measures (out of print)
Volume 3: Tests for Special Populations, 1989
Volume 4: Cognitive Aptitude Tests, 1990
Volume 5: Attitude Tests, 1991
Volume 6: Personality Tests, 1992

Keyser, Daniel J. and Richard C. Sweetland (Eds.). *Test Critiques (Volumes I-IX).* Austin, TX: PRO-ED, Inc., 1984-92.

Kramer, Jack J. and Jane Close Conoley (Eds.). *The Eleventh Mental Measurements Yearbook.* Lincoln, NE: The University of Nebraska—Lincoln, Buros Institute of Mental Measurements, 1992.

Mitchell, James V., Jr. *Tests in Print (3rd ed.).* Lincoln, NE: The University of Nebraska—Lincoln, Buros Institute for Mental Measurements. (4th edition to be published in 1994)

Sweetland, Richard C. and Daniel J. Keyser (Eds.). *Tests: A Comprehensive Reference for Assessments in Psychology, Education, and Business* (3rd ed.). Austin, TX: PRO-ED, Inc.

Articles and Books

General changes in assessment, instruction, and goals for students:

American Federation of Teachers; National Council on Measurement in Education; and National Education Association. "Standards for Teacher Competence in Educational Assessment of Students." *Educational Measurement: Issues and Practice* 9 (1990 Winter): 30-32.

Bond, Linda A. *Executive Summary of Issues and Recommendations Regarding Implementation of High School Graduation Tests.* Oak Brook, IL: North Central Regional Educational Laboratory, Regional Policy Information Center.

This paper focuses on the practical policy issues that a state should consider as it adopts a high school graduation test.

Educational Testing Service. *Policy Information Report - Testing in America's Schools.* Princeton, NJ: ETS, Policy Information Center, 1994.

Describes the types and prevalency of assessments being undertaken by states.

Gardner, Howard. "Educating for Understanding." *The American School Board Journal,* July 1993, pp. 21-24.

A good overview of the changing goals we have for students.

Herman, Joan L., Pamela R. Aschbacher, and Lynn Winters. *A Practical Guide to Alternative Assessment.* Alexandria, VA: Association for Supervision and Curriculum Development, 1992.

Hymes, Donald. "The Changing Face of Testing and Assessment." *AASA*, 1991, (703) 528-0700

A good overview of changes in assessment and the reasons for them.

McDaid, Janet L., and Donna G. Davis. "Using Assessment Results Wisely." *Thrust for Educational Leadership,* October 1991, pp. 34-36.

Considerations when designing an assessment system, including the roles and responsibilities of various interested parties.

Murray-Ward, Mildred, and Annie W. Ward. ***Assessment for All Students: Where Have We Been and Where Are We Going?*** **California Lutheran University, Thousand Oaks, CA.**

Presents a brief history of testing and assessment beginning with civil service testing in China in 2200 BC and ending with current debates on assessment. The authors try to point out how lessons from the past can help us in current endeavors.

Newmann, Fred. "Linking Restructuring to Authentic Student Achievement." ***Phi Delta Kappan,*** **February 1991, p. 458.**

Makes the point that restructuring has to be goal oriented. The first question to ask is "What changes do I want to see in students?" and then to ask "How will I restructure to achieve these student goals?"

Paul, Richard W. and Gerald M. Nosich. "A Model for the National Assessment of Higher Order Thinking." Located in: A.J.A. Binker (Ed.) ***Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World, Revised Edition,*** **1992, pp. 78-123. Available from: The Foundation for Critical Thinking, 4655 Sonoma Mountain Rd., Santa Rosa, CA 95404, (707) 546-4926.**

This paper provides a model for the national assessment of higher order thinking. The first main section of the paper states 21 criteria for higher-order thinking assessment.

Paul, Richard W. and Gerald Nosich. "Using Intellectual Standards to Assess Student Reasoning." In A.J.A. Binker (Ed.) ***Critical Thinking: What Every Person Needs to Survive in a Rapidly Changing World,*** **Revised Second Edition, 1992. Foundation for Critical Thinking, 4655 Sonoma Mountain Rd., Santa Rosa, CA 95404, (707) 546-4926.**

Describes components of reasoning.

Stiggins, Richard J. ***Student-Centered Classroom Assessment.*** **New York: Macmillan College Publishing Company, 1994.**

A textbook on the use of different types of assessments (selected response, essay, performance, and personal communication) to evaluate student attainment of knowledge, reasoning, skill, product and affective outcomes. The focus is on classroom assessment.

UCLA, Center for the Study of Evaluation and The National Center for Research on Evaluation, Standards, and Student Testing. *Evaluation Comment*, Winter 1994. Los Angeles: UCLA, CSE/CRESST, Graduate School of Education.

This issue deals mainly with the proceedings of the 1993 CRESST Conference, "Assessment Questions: Equity Answers."

UCLA, Center for the Student of Evaluation and The National Center for Research on Evaluation, Standards, and Student Testing. *Evaluation Comment*, Spring 1993. Los Angeles: UCLA, CSE/CRESST, Graduate School of Education.

This issue deals mainly with the proceedings of the 1992 CRESST Conference, "What Works in Performance Assessment?"

Literacy assessment, instruction and goals for students:

Calfee, Robert. "The Teacher's Role in Using Assessment to Improve Learning." *Assessment in the Service of Learning*, ETS, 1988, (609) 734-5686.

Describes current and past notions of reading instruction and how assessment relates to these notions.

Stayter, Francine. "Evaluating the Teaching and Learning of Literacy." In T. Shanahan (Ed.), *Reading and Writing Together: New Perspectives for the Classroom*, Norwood, MA: Christopher-Gordon Publisher Inc., 1990.

Describes how, if done correctly, assessment can be a tool for learning and not just a tool for monitoring.

Mathematics assessment, instruction, and goals for students:

Assessment Standards for School Mathematics (working draft) October 1993. NCTM, 1906 Association Dr., Reston, VA 22091, (703) 620-9840.

Clarke, David. *Assessment Alternatives in Mathematics.* Curriculum Corporation, St. Nicholas Pl., 141 Rathdowne St., Carlton, Vic 3053, Australia.

A set of materials including a teachers' resource booklet, a booklet of blackline masters, and a booklet of linked lessons to support assessment alternatives.

Commission on Standards for School Mathematics. ***Curriculum and Evaluation Standards for School Mathematics.*** **National Council of Teachers of Mathematics, 1906 Association Dr., Reston, VA 22091.**

Considered the "bible" for curriculum standards in mathematics.

Lacampagne, Carole B. ***Transforming Ideas for Teaching and Learning Mathematics.*** **US Department of Education, 555 New Jersey Ave. NW, Washington DC 20208, (202) 783-3238.**

An excellent overview of what is going on in mathematics goals for students, instruction and assessment. This would be good for parents, school administrators, and others needing a quick overview.

Mathematical Sciences Education Board. ***Measuring What Counts: A Policy Brief.*** **National Academy Press, Washington DC, 1993.**

A good statement of the position that assessment should be a tool for instruction, and not just for monitoring student progress.

Regional Educational Laboratories. ***EdTalk: What We Know About Mathematics Teaching and Learning.*** **Council for Educational Development and Research, 200 L Street NW, Suite 601, Washington, DC 20036, (202) 223-1593.**

An excellent summary of current thinking about goals for students, instruction, assessment, and the role of the community.

Zarinnia, E. Ann, and Thomas Romberg. "A Framework for the California Assessment Program to Report Student's Achievement in Mathematics." Thomas Romberg (Ed.), ***Mathematics Assessment and Evaluation: Imperatives for Mathematics Educators.*** **State University of New York Press, State University Plaza, Albany, NY 12246, 1992.**

This paper takes the position that assessment affects instruction, and therefore, regardless of the other purposes for the assessment, the instructional implications of our assessments must be taken into account.

Science assessment, instruction, and goals for students:

McKinney, Kay. ***Improving Math and Science Teaching.*** **US Government Printing Office, P.O. Box 371954, Pittsburgh, PA 15250-7954, 1993. Stock No. 065-000-00553-1.**

Summarizes the presentations at the Secretary's Second Conference of Math and Science concerning the reasons why science and mathematics instruction must change, the goals for such instruction, and implications for materials, curriculum, assessment and teacher preparation.

National Science Teachers Association. ***Scope, Sequence and Coordination of Secondary School Science, Volume 1: The Content Core, A Guide for Curriculum Designers.*** **The National Science Teachers Association, Special Publications Department, 1742 Connecticut Ave. NW, Washington, DC 20009, (202) 328-5800.**

One of two books currently considered the best statement of outcomes for students.

Regional Educational Laboratories. ***EdTalk: What We Know About Science Teaching and Learning.*** **Council for Educational Development and Research, 200 L Street NW, Suite 601, Washington, DC 20036, (202) 223-1593.**

An excellent summary of current thinking about why science is important for all citizens, goals for students, instruction, assessment, what teachers need, and the role of parents.

Rutherford, F. James, and Andrew Ahlgrem. ***Science for All Americans.*** **Oxford University Press, Inc., 200 Madison Ave., New York, NY 10016, (800) 334-4249.**

One of two books currently considered the best statement of outcomes for students.

Writing assessment, instruction and goals for students:

Spandel, Vicki, and Richard Stiggins. ***Creating Writers.*** **New York: Longman, 1989.**

How to improve student writing ability by teaching students to be self-assessors of writing.

The change process

Changing one's approach to assessment is the same as any other change—it takes time, resources, commitment and training. These articles are included as good summaries of what it takes to make meaningful change in educational practice.

Fitzpatrick, Kathleen. "Restructuring to Achieve Outcomes of Significance for All Students." ***Educational Leadership*** **49, May 1992, pp. 18-22.**

Describes one district's experience with restructuring.

Ledell, Marjorie, and Arleen Arnsparger. ***How to Deal With Community Criticism of School Change,*** **1993. ASCD, 1250 N. Pitt St., Alexandria, VA 22314, (703) 549-9110, Stock No. 611-93012.**

This thin publication has sections dealing with communication, building support, working with the media, and what to do when you're under attack.

Muncey, Donna, and Patrick McQuillan. "Preliminary Findings from a Five-Year Study of the Coalition of Essential Schools." ***Phi Delta Kappan*** **74, February 1993, pp. 486-494.**

True restructuring is very hard, time-consuming, and takes training.

Phi Delta Kappan. "Is Your School Ready for Alternative Assessment?" ***Phi Delta Kappan,*** **February 1993, pp. 455-456.**

Lists ten conditions that are important indicators of a school's readiness to embark on any new alternative assessment effort.

Regional Laboratory for Educational Improvement of the Northeast and Islands, The, 300 Brickstone Square, Suite 950, Andover, MA 01810

Publishes *Facilitating Systemic Change in Science and Mathematics Education: A Toolkit for Professional Developers*, collaboratively developed by the 10 regional educational laboratories.

Shoemaker, Betty. ***Implementation of Oregon's Educational Act for the Twenty-first Century.*** **Eugene School District, 200 N. Monroe St., Eugene, OR 97402-4295, 1993.**

An excellent summary of the elements important for successful change.

Annotated Bibliographies

Arter, Judith A. *Assessment Alternatives in Science, (updated regularly).* Portland, OR: Northwest Regional Educational Laboratory.

Arter, Judith A. *Bibliography of Assessment Alternatives: Social Studies,(updated regularly).* Portland, OR: Northwest Regional Educational Laboratory.

Arter, Judith A. *Math Assessment Alternatives, (updated regularly).* Portland, OR: Northwest Regional Educational Laboratory.

Arter, Judith A. *Portfolio Resources, (updated regularly).* Portland, OR: Northwest Regional Educational Laboratory.

Arter, Judith A. *Reading Assessment Alternatives, (updated regularly).* Portland, OR: Northwest Regional Educational Laboratory.

VIDEOTAPE RESOURCES

Alternatives for Measuring Performance. **IRI/Skylight Publishing, Inc., 200 E. Wood St., Suite 250, Palatine, IL 60067, (800) 348-4474, ext. 220.**

On November 21, 1991, the North Central Regional Educational Laboratory broadcast a national teleconference on assessment to sites throughout the United States. This videotape (60 minutes) of the teleconference profiles two schools utilizing innovative assessments. A discussion by a distinguished panel of assessment experts follows. The videotape is accompanied by a thirty-six page guidebook which describes the characteristics of effective assessments, criteria for valid performance assessments, and recommended resources.

Arts PROPEL Video Handbook, The. **Harvard Project Zero, Longfellow Hall, 13 Appian Way, Cambridge, MA 02138, (800) 235-2132.**

Arts PROPEL was initiated to develop nontraditional models of assessment appropriate for students engaged in the visual arts, music, and imaginative writing. This video (65 minutes) accompanies the Arts PROPEL Handbook for Visual Arts and provides an overview to the Arts PROPEL project. The program includes sections on PROPEL's guiding principles, students engaged in long-term visual arts projects, and the use of portfolios as vehicles for communication and reflection. Teacher testimonials lend authenticity and practicality to the project.

Assessment and Technology. **The Center for Technology in Education, Bank Street College of Education, 610 W. 112th St., New York, NY 10025.**

This video was produced in conjunction with a research project to determine how video and computer technology can best be used to assess learning in the science and mathematics. The video begins with a description of one school's use of technology-based science assessments. The program then presents four examples of a performance assessment in which eleventh graders work in teams to design and construct motorized devices to demonstrate their understanding of important concepts in applied physics. Printed scoring criteria are provided and viewers are encouraged to evaluate the student performances as they view the video.

California Learning Record, The. **California Department of Education, Bureau of Publications, Sales Unit, PO Box 271, Sacramento, CA 95812, (916) 445-1260.**

This video was produced by the California Department of Education to introduce the California Learning Record (CLR), a developmentally appropriate portfolio assessment process for young children. Filmed in schools with Chapter 1 programs in the San Diego Unified School District, the program shows parents, teachers, and students making contributions to the CLR.

Classroom Assessment Workshops. **IOX, 5301 Beethoven St., Suite 109, Los Angeles, CA 90066, (310) 822-3275.**

This is a series of 14 interactive videotapes dealing with classroom assessment. The tapes were developed by Northwest Regional Educational Laboratory and have running times from one hour to all-day. The series includes: *A Status Report on Classroom Assessment; Understanding the Meaning and Importance of Quality Classroom Assessment; Developing Assessments Based on Observation and Judgment; Paper and Pencil Test Development; Measuring Thinking in the Classroom; Assessing Reading Proficiency; Writing Assessment: Issues and Answers; Writing Assessment: Training in Analytical Scoring; Developing Sound Grading Practices; Understanding Standardized Tests; Assessing Mathematical Power; Assessment in the Science Classroom; Using Portfolios in Assessment and Instruction;* and *Facing the Challenges of a New Era of Educational Assessment.*

In the Classroom. **IOX, 5301 Beethoven St., Suite 109, Los Angeles, CA 90066, (310) 822-3275.**

This new series currently includes two videos: *Writing from the Inside Out: Revising for Quality* and *Putting Portfolio Stories to Work.* Each title has a 20-minute video, workshop facilitator's guide, and complete 125 page textbook. The videos are designed to run continuously, or stop at strategic points for workshop discussion and activities. The facilitator's guides answer many questions about presenting workshops and provide helpful trainer's tips.

Mathematics Assessment: Alternative Approaches. **National Council of Teachers of Mathematics, 1906 Association Dr., Reston, VA 22091, (800) 235-7566.**

This video (71 minutes) was produced in cooperation with South Carolina Television as a companion to the book, *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions* by Jean Stenmark. Through classroom examples and expert commentary, the program explores teachers' implementations of a variety of assessment methods in their classrooms. A viewer's guide is included.

Performance Assessment in the Classroom. **The Video Journal of Education, 549 W 3560 S, Salt Lake City, UT 84115, (800) 572-1153.**

This consists of two tapes: *Part 1: Performance-Based Assessment and Instruction* (29 minutes); and *Part 2: Creating Performance Tasks* (31 minutes). These two videotapes are designed to introduce teachers, administrators, and parents to the key principles of performance-based assessment and instruction.

The program invites the viewer to consider ten key questions related to the design and use of classroom performance tasks. Numerous classroom examples and comments by teachers make this a practical and fast-paced resource for staff development.

Portfolio Assessment in the Classroom. **The Video Journal of Education, 549 W 3560 S, Salt Lake City, UT 84115, (800) 572-1153.**

This consists of two tapes: *Program 1: Reflections of Learning* (29 minutes); and *Program 2: Utilizing Portfolios* (29 minutes). These two videotapes feature Bena Kallick, a noted expert on portfolio assessment. The program emphasizes the use of portfolios to document learning and stimulate self-evaluation by students. A variety of classroom examples illustrate the use of portfolio assessment in different subject areas and grade levels.

Redesigning Assessment. **Association for Supervision and Curriculum Development, 1250 N. Pitt St., Alexandria, VA 22314, (703) 549-9110.**

This series contains three tapes: Tape 1—*Redesigning Assessment: Introduction* (24 minutes); Tape 2—*Redesigning Assessment: Performance Assessment* (32 minutes); and Tape 3—*Redesigning Assessment: Portfolios* (40 minutes). It features classroom illustrations of performance assessment in action with comments by students, teachers, administrators, and national experts. Each videotape includes a leader's guide offering suggestions for using the videos as part of a workshop for professional staff, parents, and boards of education.

These tapes are available for rental or purchase.

Standards NOT Standardization. **Center for Learning, Assessment, and School Structure, 39 Main St., Geneseo, NY 14454, (716) 243-5500.**

This series contains four volumes: *Volume I—Re-thinking Assessment: Provocations and Issues; Volume II—Re-thinking Assessment: The District; Volume III—Re-thinking Assessment: The Classroom; and Volume IV—Re-thinking Assessment: The School.* This series serves as the centerpiece for a series of seminars on raising the standards of student performance through attention to "authentic" assessments. Each volume contains two videotapes and a comprehensive manual providing detailed suggestions for seminar leaders.

Student Growth: Directions for Assessing and Communicating Learning. **Edmonton Public Schools, Center for Education, 1 Kingsway, Edmonton, Alberta, Canada T5H 4G9.**

This video (16 minutes) describes the use of a portfolio system in Edmonton Public Schools for assessing six essential learning outcomes identified by the district. The video shows examples of portfolio review conferences with staff, students, and parents.

Teacher TV - Episode 11: Alternative Assessment. **NEA Professional Library, PO Box 509, West Haven, CT 06516, (800) 229-4200.**

Co-produced by the National Education Association (NEA) and the Learning Channel, this video (22 minutes) contains vignettes of teachers using "alternative" methods for assessing student learning. One example describes the use of a computer and scanner for assembling a digital portfolio at the elementary level. A second vignette illustrates how a teacher involved her students in peer evaluation as part of a science fair project.

Teaching and Testing in Maryland Today: Education for the 21st Century. **Maryland Assessment Consortium, 115 E. Church St., Frederick, MD 21701, (301) 694-1800.**

This video (14 minutes) was developed for use with school staffs, parents, community members, and boards of education. It provides an overview of the new role of performance-based instruction and assessment in the context of Maryland's outcome-based educational reforms.